

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2001, 23:05:42 ; Search time 7860.64 Seconds
(without alignments)
1923.232 Million cell updates/sec

Title: US-09-477-392-1
Perfect score: 2954
Sequence: 1 cgcgttgcacgtctc.....aaaaaaaaaaaaaaaaaaaaa 2954

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues
tal number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pl1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_cm:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vl:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: em_da1:*
34: em_da2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_vil1:*
72: gb_vil2:*
73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2954	100.0	2954	37 AF201303	AF201303 Homo sapi
2	2729	92.4	132150	9 AC005586	AC005586 Homo sapi
3	2719.4	92.1	2878	37 AK025356	AK025356 Homo sapi
4	2654.6	89.9	2878	77 HSA245553	AJ245553 Homo sapi
5	1256.4	42.5	179225	58 AC073111	AC073111 Homo sapi
6	1194.8	40.4	201458	40 AC015887	AC015887 Mus muscu
7	719	24.3	759	78 HSM802596	AL162065 Homo sapi
8	315.8	10.7	6045	8 AB002324	AB002324 Human mRN
9	315.8	10.7	185306	39 AC013570	AC013570 Homo sapi
10	315.4	10.7	239566	77 HSMJ03147	AJ003147 Homo sapi
11	315.2	10.7	2382	37 AK023017	AK023017 Homo sapi
12	310.4	10.5	130067	9 AC007228	AC007228 Homo sapi
13	304.6	10.3	2094	11 MM2EP29	X55126 M.musculus
14	299.8	10.1	183525	39 AC012313	AC012313 Homo sapi
15	283.6	9.6	230534	65 AC079575	AC079575 Mus muscu
16	282.8	9.6	128361	85 AC004877	AC004877 Homo sapi
17	281	9.5	203100	55 AC060772	AC060772 Mus muscu
18	279.4	9.5	295	78 HSG5	X82192 H.sapiens E
19	278.4	9.4	2311	78 HS2NF74A3	AF072567 Homo sapi
20	278.4	9.4	2416	78 HS2NF741	X71623 H.sapiens Z
21	278.4	9.4	2416	79 G27154	G27154 SHCC-31580

C	22	278.4	9.4	3759	78	HSZNE742	X92715 H. sapiens m
C	23	278.4	9.4	183617	9	AC007731	AC007731 Homo sapi
C	24	278.4	9.4	193592	8	AC005500	AC005500 Homo sapi
C	25	277.8	9.4	4582	11	MUSLDM24X	M98502 Mus musculus
C	26	277.6	9.4	2587	10	AF055077	AF055077 Homo sapi
C	27	277.6	9.4	2678	85	HUM4Z61	M82827 Human zinc
C	28	277.6	9.4	2905	10	AF055078	AF055078 Homo sapi
C	29	277.6	9.4	12456	37	AF161886	AF161886 Homo sapi
C	30	277.6	9.4	40616	41	AC016630	AC016630 Homo sapi
C	31	277.6	9.4	44210	41	AC016629	AC016629 Homo sapi
C	32	275.6	9.3	2869	11	AB007407	AB007407 Mus muscu
C	33	274.2	9.3	113247	65	AC0079506	AC0079506 Mus muscu
C	34	273.8	9.3	1554	37	AF269249	AF269249 Homo sapi
C	35	273.8	9.3	130067	9	AC0071228	AC0071228 Homo sapi
C	36	270.2	9.1	210923	39	AC013273	AC013273 Homo sapi
C	37	263.8	8.9	138278	84	HSY99615	AL109966 Homo sapi
C	38	260.6	8.8	166803	55	AC048382	AC048382 Homo sapi
C	39	260.6	8.8	194486	41	AC016771	AC016771 Homo sapi
C	40	250.2	8.5	177635	38	AC010624	AC010624 Homo sapi
C	41	246.4	8.3	2624	8	AB046644	AB046644 Macaca fa
C	42	245.4	8.3	1800	4	GGCKR1	X15558 Chicken ckr
C	43	245	8.3	179215	58	AC073422	AC073422 Homo sapi
C	44	242.2	8.2	139408	65	AC079555	AC079555 Mus muscu
C	45	238.2	8.1	175999	49	AC021154	AC021154 Homo sapi

ALIGNMENTS

X927215 H. sapiens m
 AC007721 Homo sapi
 AC005500 Homo sapi
 W85052 Mus muscul
 M850507 Mus muscul
 M58297 Human zinc
 AC050578 Homo sapi
 AF161886 Homo sapi
 AC016630 Homo sapi
 AC016629 Homo sapi
 AB007407 Mus muscu
 AC079506 Mus muscu
 AF265249 Homo sapi
 AC007228 Homo sapi
 AC013373 Homo sapi
 AL109966 Homo sapi
 AC048382 Homo sapi
 AC016771 Homo sapi
 AC010624 Homo sapi
 AB046644 Macaca fa
 AL15538 Chicken chr
 AC073422 Homo sapi
 AC075965 Mus muscu
 AC021154 Homo sapi

RESULT	1
LOCUS	AF201303
DEFINITION	Homo sapiens dhfr oribeta-binding protein RIP60 mRNA, complete cds.
ACCESSION	AF201303
VERSION	AF201303.1 GI:6716713
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2954) Houchens,C.R., Montigny,W., Zeltser,L., Dalley,L., Gilbert,J.M. and Heintz,N.H. The dhfr oribeta-binding protein RIP60 contains 15 zinc fingers: DNA binding and looping by the central three fingers and an associated proline-rich region Nucleic Acids Res. 28 (2), 570-581 (2000) 10606657 2 (bases 1 to 2954) Houchens,C.R., Gilbert,J.M. and Heintz,N.H. .Direct Submission Submitted (02-NOV-1999) Pathology, University of Vermont, Soule Medical Alumni Building, Burlington, VT 05405, USA Location/Qualifiers 1..2954
TITLE	
JOURNAL	
PUBMED	
ERENCE	
UTHOES	
TITLE	
JOURNAL	
FEATURES	
source	

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HeLa"
130.1833
/note="contains 15 C2H2-Kruppel-like zinc fingers"
/codon_start=1
/product="Ohf oribeta-binding protein R160"
/protein_id="AAF26712.1"
/db_xref="GI:6716714"
/translation="MLERRCKGLPVLAMGLAORILSGSPQESPTLGKESRLROOGTS
VAOSGAOPGABRACACRRHFGWALWLTTRTCOARLPYCEGCGEGRFAFLH
HROVHAATPDLGFACHLGSGESFGWALVYLHILHAAKQPIACPCEREMRKOI
RAHLRRCRPPAPRAEPTICGSGSPFMOLVLAHKVYLAALAEPAARCKTPRPG
RYAVTAARPGSDANDRPFOCACCCRRRRHNPILAHRRVHTGEPRHOCPGCGKRTNK
PGLSSAARPTPAVPLKPAOEPPEGAPPEHDDPLEAPSLTSCDDCGSFELEFLR
PILSHRRITPIGEKPTCKCEGGRFRHNPILSHSKLHKRSGSAAPGGSFPLRA
HÖRDHSTDEPTCAEGKNFQKTHLVASHSVHGEFPAECGGRASÖÖSHLAANR
PDNAHPDPRFVCPDGGKAFRHKPYLARRHITHTGKPYVCPDGGKAFSÖKSNLVSHRR

```

[illegible]

QY 961 tatctgattcgcagccgagcatccacacccgagagaagccctacccgtgcaaaagtg 1020
 |||||
 Db 961 TATCTGATCTTCGACCCGCGCATTCACACCGCGAGAAAGCCCTACCGCTGCAAGAGTGC 1020
 QY 1021 ggcgcgcgtctccgcgcaaaacccaactgtctgtctacagcaagatltcaacagcgtc 1080
 |||||
 Db 1021 GGC GCGCGCTTCCGGGCAAAACCAACCTGCTGTCTCAGACGAAGATTCAACAGCATCC 1080
 QY 1081 gagggtctggcagcagc 1140
 |||||
 Db 1081 GAGGAGGCTGGCGCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
 QY 1141 gattcgc 1200
 |||||
 Db 1141 GAGTCCG 1200
 QY 1201 ggggc 1260
 |||||
 Db 1201 GGGGCG 1260
 QY 1261 gacgaactggcagagagcttcggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320
 |||||
 Db 1261 GACGACTGCG 1320
 QY 1321 accgggagagc 1380
 |||||
 Db 1321 ACCGGGAGAGCG 1380
 QY 1381 ctgtgtgagcactcgc 1440
 |||||
 Db 1381 CTGTGTGCG 1440
 QY 1441 cgc 1500
 |||||
 Db 1441 CG 1500
 QY 1501 cccttctgtctcgc 1560
 |||||
 Db 1501 CCTTCTGTGTCG 1560
 QY 1561 cgc 1620
 |||||
 Db 1561 CG 1620
 QY 1621 cagaatcacaactgt 1680
 |||||
 Db 1621 CAGAACTCCAACTGT 1680
 QY 1681 cgc 1740
 |||||
 Db 1681 CCGGACTGCG 1740
 QY 1741 atccgc 1800
 |||||
 Db 1741 ATCCGGGAGCGCGCGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800
 QY 1801 ctctctgc 1860
 |||||
 Db 1801 CTCTCTGCG 1860
 QY 1861 gagaagc 1920
 |||||
 Db 1861 GAGAGGCGCTGGGCTCTTCTGT 1920
 QY 1921 tggagtgaggagacatggaatcctagaggagatgagagatgagagatgagagatgagagatg 1980
 |||||
 Db 1921 TGAGTGAAGGAGACATGGAATCCTAGAGGAGATGGAAGATGGGGAGAGTGTGAGTGTG 1980
 QY 1981 ggc 2040
 |||||
 Db 1981 GGCGCTGCTAGCGAGAGAGTCAACCCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040

QY 2041 agc 2100
 |||||
 Db 2041 AGCGCGCGCTTCACCTGCTGT 2100
 QY 2101 agc 2160
 |||||
 Db 2101 AGCAGCATATCATCGCGCGCTTACAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
 QY 2161 ggc 2220
 |||||
 Db 2161 GCGCTTCATCTCTGT 2220
 QY 2221 atcatttggagagc 2280
 |||||
 Db 2221 ATCATTTTGGAGACAGCGCGGT 2280
 QY 2281 ctltgagaacgc 2340
 |||||
 Db 2281 CTTGAGAGACCTGCTGGCGCTTGT 2340
 QY 2341 aagcgaagc 2400
 |||||
 Db 2341 AAGCGAAGCGCGCTTACCT 2400
 QY 2401 acttgagagagc 2460
 |||||
 Db 2401 ACTTGAGAGAGCGCGGT 2460
 QY 2461 ccaagagcattacgc 2520
 |||||
 Db 2461 CCAAGGCAATTACCT 2520
 QY 2521 cctccgc 2580
 |||||
 Db 2521 CCTCCCTCTGCTGT 2580
 QY 2581 ttgcctcttccagagc 2640
 |||||
 Db 2581 TTGCT 2640
 QY 2641 caagctgagggagctgt 2700
 |||||
 Db 2641 CAGCTGAGGGGAAATGCTGT 2700
 QY 2701 ctctgtctgagcactgagtgagagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2760
 |||||
 Db 2701 CTCTGTGTGAGCACTTGT 2760
 QY 2761 aattgttagaattctaatgt 2820
 |||||
 Db 2761 AATTGTTAGAATTCTTAATGT 2820
 QY 2821 tataatttcaactatgt 2880
 |||||
 Db 2821 TATAATTTCATACCTATGT 2880
 QY 2881 cccctccttccaactagagataaacttggctttagaagagagagagagagagagagagagag 2940
 |||||
 Db 2881 CCCCTCTCTTTCAAACTAGAGATAAATCATTTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAG 2940
 QY 2941 aaaaaaaaaaaaaa 2954
 |||||
 Db 2941 AAAAAAAAAAAAAA 2954

RESULT 2
 AC005586/c 132150 bp DNA PRI 30-SEP-2000
 LOCUS AC005586 Homo sapiens PAC clone RP4-584D14 from 7q31-q35, complete sequence.
 DEFINITION AC005586
 ACCESSION AC005586.2 GI:8468933
 VERSION HTG.
 KEYWORDS human.
 SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 132150)

AUTHORS

Sulston, J.E. and Waterston, R.

TITLE

Toward a complete human genome sequence

MEDLINE

Genome Res. 8(11):1097-1108 (1998)

REFERENCE

2 (bases 1 to 132150)

AUTHORS

Stonking, T., Ozersky, P., Mohlmann, P. and Le, T.

TITLE

The sequence of Homo sapiens PAC clone RP4-584D14

JOURNAL

Unpublished

REFERENCE

3 (bases 1 to 132150)

AUTHORS

Waterston, R.H.

TITLE

Direct Submission

JOURNAL

Submitted (01-SEP-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

4 (bases 1 to 132150)

AUTHORS

Waterston, R.H.

TITLE

Direct Submission

JOURNAL

Submitted (12-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

5 (bases 1 to 132150)

AUTHORS

Waterston, R.

TITLE

Direct Submission

JOURNAL

Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

On Jun 12, 2000 this sequence version replaced g1:3907529.

JOURNAL

Genome Center

AUTHORS

Center: Washington University Genome Sequencing Center

TITLE

Center code: WUGSC

JOURNAL

Web site: http://genome.wustl.edu/gsc

AUTHORS

Contact: sapient@wustl.wustl.edu

TITLE

Summary Statistics

JOURNAL

Center project name: H_DJ0584D14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see

http://www.nhgri.nih.gov/DIR/CTB/CHR7, send

mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

This clone was derived from human PAC library RPI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://pacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

FEATURES

source

The clone sequenced to the right is RP5-820A21, 200 base pair overlap. Actual start of this clone is at base position 1 of RP4-584D14; actual end is at base position 131954 of RP4-584D14.

The sequence RP4-584D14 from base position 21655 to 22775 is CT rich. This region was sized with PCR from genomic DNA and the hindIII digest with band size 3643 real, and 3642 insilico. The sequence RP4-584D14 from base position 22628 to 22637 is represented by sequence derived by PCR from genomic DNA.

location/Qualifiers

1..132150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31-q35"
/clone="RP4-584D14"
/clone_11b="RPI-4"
230..315
/rpt_family="MIR"
960..1288
/rpt_family="MER4-group"
1360..1468
/rpt_family="L2"
1562..1616
/rpt_family="L2"
1733..2080
/note="similar to EST AM510685 (NID:g7148763) hc89e08.x1"
2051..2114
/rpt_family="L2"
2148..3224
/note="CpG island (%GC=70.7, o/e=0.90, #CpGs=107)"
2347..2426
/rpt_family="CG-rich"
3141..3232
/rpt_family="Alu"
3235..3368
/rpt_family="Alu"
3795..3888
/rpt_family="MIR"
4037..4199
/rpt_family="Alu"
4308..4593
/rpt_family="MER4-group"
4872..5309
/note="similar to EST A1827463 (NID:g5448134) wf29f03.x1"
4882..4909
/rpt_family="AT-rich"
5679..5773
/rpt_family="MIR"
5947..5952
/note="match to EST AA482420 (NID:g2210098) zt34f02.r1"
6148..6450
/rpt_family="Alu"
6453..6690
/rpt_family="MER4-group"
6695..7007
/rpt_family="Alu"
7008..7117
/rpt_family="MER4-group"
7180..7335
/rpt_family="MER4-group"
7615
/note="match to EST A1144211 (NID:g3666020) qb93g11.x1"
7883..8397
/note="match to EST W72943 (NID:g1383235) zds4f12.r1"
7886..8357
/note="match to EST A1144211 (NID:g3666020) qb93g11.x1"
7886..8562
/note="similar to EST AW411281 (NID:g6936822) fh1ld05.y1"
7886..8403
/note="similar to EST AW246540 (NID:g6589533)"
7886..8451
/note="match to EST AA129215 (NID:g1689084) zn84b02.s1"

Db	9326	ACGGCCCGGATCGGCGCTTGTTGTGTCCGACGTGGCGCAAGGCGCTTCGGCCACAAACCT	9267
Qy	1547	acctgagcgagcaacgagcgatccacaacgagcgagaagccctacgtctctgcccagctcg	1606
Db	9266	ACCTGGGGGGCCACCGGGCGGATCCACACCGGGAGAAAGCCCTACGTCCTCCCGACTGG	9207
Qy	1607	gcaagcccttaagcagaagtccaaacctgtgtgcagccgagcgatccacagagggcgagc	1666
Db	9206	GCAAGCGCTTACGCCAATAAGTCCAACTGGTGTGCAACCGCGGCATCCACAGGGCGAGC	9144
Qy	1667	ggccctacgectgtcccgactcgagcccgagcttcaagcagaagttcaactcatcacc	1726
Db	9146	GGCCCTACGGCTGTGCCAGTGGGACCGCGAGCTTCAGCCAGAAGTCCAACTCATCACCC	9087
Qy	1727	accgcaagagccacatcccggaacgagcgccctctgcgtgtccatctgtgacagacctcg	1786
Db	9086	ACCCGAAAGGCCAATCCGGAGCGGCCCTTGCTGTGTGCATCTGTGGCCGAGCCTTCG	9027
Qy	1787	acgacgaggaagagatctctctgcccacacagaagaacagatctcgaacggtgagcgag	1848
Db	9026	ACGACGAGAGAGACTCTGTGGCCACACAGAAAGACAGTGTGTGAGACGGTGGGGCGG	8967
Qy	1847	ggcgtgtgtgctgtgagagaaggtctgtggttccctgtgtgtgagtcgagtcgagtcgaggt	1906
Db	8966	GCCGTGTGTGCTGTGAAGAGGAGCTGGGGCTTCCTGTGTGTGGAGTCCACTGTGGCTGGGG	8907
Qy	1907	tgccgtccctagtgtgtgagtaggaggaacatgggaatcccttagagggagatagaagatgagcg	1966
Db	8906	TGCTCGCTAOTGCTGGAGTATGGGGACAAATGGGAATCTTAGAGGGATGGAAGACGGGG	8847
Qy	1967	gaatgaagcttgggtgtgagccctgtctagcgaagagagttcaaccccggtgtgcagggaaaccaac	2026
Db	8846	GAGTAGAGCTGGGTGGGCCCTGCTAGCAGAGAGAGTCAACCCGGTGGCCAGGGAACCCAC	8787
Qy	2027	ttccaagcagcagggagacgcgcgcctccaagctgagtggtgtgtgtaaggtccgttcctgactgc	2086
Db	8786	TTCCAAAGGCGAGGAGCCCGGCGCTCCACGCTGGTGTGTGTAAGGCTCGTCTGTAATGCC	8727
Qy	2087	ctgtgcccctggaaaaagagaata-catccgccccttaagccctctgtgctagaagagagcc	2145
Db	8726	CTGTGCCCTGGAAAAAGCAGCAATATGCAATCGGCCCTTAGAGGCCCTGTGATGAGAGCC	8667
Qy	2146	accagttggaaggaag--ccctcatcctctgtgtatataagccttaatgcctctgctctt	2204
Db	8666	ACCGATGGAAAAGAAAGCCCTCATCTCTGTGATTAAGCCTTAATGCCCTGTGCTTTT	8607
Qy	2205	acttaagtttaact--aaatatcttttgaagcagcgctgtgtatagatcctgttaatgaagt	2263
Db	8606	ACTGTAAAGTTACTTAAGATCATTTTGGGAAGCAGGGGTGTGATGAGTCTGTAAATGAAATG	8547
Qy	2264	ctctgtgcttagatacagctcttgaagaaactgtgcctgtttagaacaaacttggcctt	2323
Db	8546	CTCTGGGCTAGATATCACCTTGTTGGAAACCTGTGCTGGCTGTATAGACGACTTGGGGCTTTT	8487
Qy	2324	gcacgacagcaagagtgaagcgaagcaacttaactccttccctcccaactggcc	2387
Db	8486	GCACGACACAAAGGTTAAAGCGAACCTCTTAACCTCTCCCTTCCCTCCACCTGGCC	8427
Qy	2384	ccgtgtatagcaaccagaacttggagaaccgctgtctgtlataacttcaatccctcttc	2443
Db	8426	CTGTGATAGGACCAACAGACTTGGAGAGACCGCTGCTGTATATCTTCCATTCCTTCC	8367
Qy	2444	ttcccaagagcagatctcccaagcaattactcctgtgtctgtctcgcttatctgtgcg	2503
Db	8366	TTTCCCAAAGAGCAATCCCAAAGCAATTTAATCTCTTGCTGTGCTGTATCTATCTGTGCC	8307
Qy	2504	cctccacagcctgtgagaccccccctgtgctgtacagcgaactgttccaaagcctctgtctg	2563
Db	8306	CTCTCCAGCGCTAGAGACCTCCCTGTGCTGTACAGCAACGTGTGTCCAGGCTTGTGTG	8247
Qy	2564	aaacgcgagccctctctgtctcttccaagctcaagatcaagatcaoggaagactgtcg	2623

Db	8246	AACGCGAGGCCCTCCTTGGTCCTTCCAGACTCAGATGTCAOGGGAAGACTGGCG	8187
Oy	2624	cattgtgatgagggccacgcctgaagggaagtgtcgttgatlttc- ttctccatttc	2682
Db	8186	CATTGGTGATGAGAGGGGCACCTAAGGGGAAAGTTGGTGGAATTTTCCTTTCATTTC	8127
Oy	2683	tagcatat-taacacctgcctcctgccttgcgaacttagtgatgaagaacttcgcaccctc	2741
Db	8126	TAGCATTTGAGACACCTGGCCTCTGCTTAGACATTATGATGAGAAGACTTCOGCACCTCC	8067
Oy	2742	tgaagccctggatgatctcaattgtttagaaatctcaattgtttagaaitcctcctataa	2801
Db	8066	TGAGCCCCCTGATGATTCATAATGTTAGAATTCATAATGTTAGAAAACCTTCCTTAATA	8007
Oy	2802	tgaatgaattctcttctccaataattctaacattggsccttgtctgtctctcgagac	2861
Db	8006	TGAATGAATTTCTGCTTTCCTATATATTTCTACTATTGGCCCTTGTCGTCTCTGGAC	7947
Oy	2862	taaacagaacaacattaccctcccttccaactagaagaataacattgtgtttaga	2921
Db	7946	TAAACAACAACACATTACCCTCCTTTTCAAACATAGAAATTAAGATTGGCTTTAGA	7887
Oy	2922	a 2922	
Db	7886	A 7886	
RESULT	3		
LOCUS	AK025356		
DEFINITION	AK025356	2878 bp mRNA	PRT 29-SEP-2000
ACCESSION	AK025356	Homo sapiens cDNA: FLJ21703 fis, clone COL09895, highly similar to HSA245553 Homo sapiens mRNA for AP4 zinc finger protein.	
VERSION	AK025356		
KEYWORDS	AK025356.1 GI:10437856	oligo capping; fis (full insert sequence).	
SOURCE	Homo sapiens colon cDNA to mRNA, clone_lib.COL clone:COL09895.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (sites)		
TITLE	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
JOURNAL	NEDO human cDNA sequencing project unpublished (2000) 2 (bases 1 to 2878)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan (E-mail:cdna@ems.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDN library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
FEATURES	Location/Qualifiers		
SOURCE	1..2878		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="COL09895"		
	/clone_lib="COL"		
	/issue_type="colon"		
	/note="Cloning vector pME18SFL3"		
	1..2878		
	/note="highly similar to HSA245553 Homo sapiens mRNA for AP4 zinc finger protein"		

BASE COUNT 560 a 974 c 826 g 518 t
ORIGIN

Query Match 92.1%; Score 2719.4; DB 37; Length 2878;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 2806; Conservative 0; Mismatches 31; Indels 8; Gaps 6;

QY 115 gcaagagaaagaaacgaatgagaaacgtcgttgaagggccccctggcaccatgagcttgcc 174
DB 13 GCAAGAGAAAGAAACGATGCTGGAACGCTGTGAGAGGGCCCCCTGGCCATGAGGCCGCC 72

QY 175 caagcccgactcctctcttgagccctcccaagagtcacccacagaccctgggaaagatcc 234
DB 73 CAGCCCGACTCTCTTGTGGGCCCTCCAGAGTCAACCCAGACCCCTGGGGAAAGATCC 132

QY 235 cgcgggctgagaaacaaagcaagtcagtgagccagtcgtgtgccaagcccccaagcag 294
DB 133 CGCGGGCTGAGCAACAAAGCAGCTCAGTGGCCAGTCTGTGCCCCAAGCCCCAGCAGG 192

QY 295 gccatcgtgtgccaactgagaaagcattccctgagtggtgtgtgtgtgtgtgtgtgtgt 354
DB 193 GCCCATGCTGTGCTCCACTCAAGGCACTTCCCTGGGTGGGTGCTGTGTGTGTGTGTGT 252

QY 355 acccgccgctgagcagagcccgctgacctgacctgacctgacctgacctgacctgacctg 414
DB 253 ACCCGCCGCTGAGCGCCCGGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 312

QY 415 catgccccctcttaagcaactgacacgacgacgacgacgacgacgacgacgacgacgac 474
DB 313 CATGCTCCCTCTTGTAGACTGACACCGCAGGTCATGCTGCTGCTGCTGCTGCTGCTGCT 372

QY 475 ttgtgctgcaactctgtgtgagcagagcttcagaggtcgtgtgtgtgtgtgtgtgtgtgt 534
DB 373 TTGTGCTGCTCACTGTGTGGGAGAGCTTCCGAGGCTGGGTGGGTGCTGTGTGTGTGTGT 432

QY 535 ctggcccatcagctgcaaaagcaaccatcgtctgtctccaaatgacgagagacgtctgtg 594
DB 433 CGGGCCCATTCAGCTGCAAAAGCGGCCCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 492

QY 595 cgaagaaagacgtctgagctcatctgtgagcggtgacacccctcccgcccgagagcccg 654
DB 493 CGAGAAAGACGCTTGAGCTCATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 552

QY 655 ccctcataatgagcaactgtgagcagagcttggccagtgaggagacagctagtgtgtgtgt 714
DB 553 CCTTCAATATGCGGCAACTGTGGCCGGAGCTTGTGCCAGTGGGACAGCTAGTTGCCAC 612

QY 715 aagcgggtgacgtaagctgagggcccttgagagagcgacgacgaaggtctgtgagcccg 774
DB 613 AAGCGGTGCACTAGCGGAGGCGGCTGTGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 669

QY 775 cccaggggcccgcgcgctgacgcccccgcccccgcccccgcccccgcccccgcccccg 834
DB 670 CCCAGGGGCGCCCGCGGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 729

QY 835 ttccagtgtgctgctgtgtgtgcaagcgtctccgcaacagcccaactgtatcgtctacgc 894
DB 730 TTCTCAGTGTCTGT 789

QY 895 cgcgtgacacgagcgagggcgccccacacagtgccccagtgccccagtgccccagtgcccc 954
DB 790 CGCGTGCACACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 849

QY 955 aagcccatctgacttgcacacgagcattcacacacgagcagagaccccttaccctgtgcaaa 1014
DB 850 AAGCCCATCTGACTTGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 909

QY 1015 gagtgtgagccgctgtccgcaacaaacccacacacacacacacacacacacacacacac 1074
DB 910 GAGTGTGCGCGCGGCTTCCGCGCAAAACCACTGCTGTCTCAACAGGATTTCAACAG 969

QY 1075 cgttcagaggtgtgagccagcgcccccgcccccgcccccgcccccgcccccgcccccg 1134

DB 970 CGATCCGAGGGGTGCGGCCAGGCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1029

QY 1135 ccccaagagatccgctgagcagagcccaaccccgagctacctgtgaaacccgagagagcg 1194
DB 1030 CCCAGAGAGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1089

QY 1195 ccgcgaagggcccccgcaagagcaccgcaagaccgagtgaaagccccctccctctac 1254
DB 1090 CCGCAGAGGGGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1149

QY 1255 agctgtgacagactgtgagagagcttccgctgtgagagcttctgtgagcccaacagcag 1314
DB 1150 AGCTGCGACGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1209

QY 1315 caagcaacccgggagcgcccttcaactgagcagagtgagagagaaacttgcgaagaag 1374
DB 1210 CAGCAGACCGGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1269

QY 1375 acgacactgtgtgagacactgagcagtgactcagagagagagagagagagagagagag 1434
DB 1270 ACGCAGCTGTGTGGCGCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1329

QY 1435 tgcggccgctctctcccaagagcagcagcagcagcagcagcagcagcagcagcagcagc 1494
DB 1330 TGGGCGCGCGCTTCTCTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1389

QY 1495 gatcgccctctgt 1554
DB 1390 GATCGGCGCTTCTGT 1449

QY 1555 cggacacggcgatcccaacacgagagaaagccctacgtctgtgtgtgtgtgtgtgtgtgt 1614
DB 1450 GCGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1509

QY 1615 ttcaagcaagatcccaacctgt 1674
DB 1510 TTCAAGCAAGAGTCCAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1529

QY 1675 gctgtccagactgtgagcagcagcttcaagcagaagtcacacatcaacccacgcaag 1734
DB 1570 GCTGTCCGCACTGT 1629

QY 1735 agccacatccggagagcgccctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1794
DB 1630 AGCCACATCCGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1689

QY 1795 gaagactctctgtgcccacagaagaagcacgatlcttgagacgtgtgagcgagcgctgtgt 1854
DB 1690 GAGAGACTCTCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1749

QY 1855 ggtctagaagaggtgtggttctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1914
DB 1750 GGCTGAGAGAGAGGCTGTGGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1809

QY 1915 taagtgtgaggtgtgagaaatctctagaagagagagagagagagagagagagagagag 1974
DB 1810 TAAGTGTGAGGTAGGAGACATGGGAATCTTAGAGGGGAGTGGAGAGCGGCGGAGTAGC 1869

QY 1975 ttgtgtgacctgtgag 2034
DB 1870 TGGGTGGGCGCTGTACGAG 1929

QY 2035 gcaaggaagcgcgcctcagactgt 2094
DB 1930 GCAAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1989

QY 2095 ttgaaagagagaaata-cattcgcccttaagagccctctgtgtgtgtgtgtgtgtgtgtgtgt 2153
DB 1990 TGGAAAGAGAGCAATTAACATCCGCCCTTGAAGGCCCTGTGATGAGAGAGAGAGAGAGAG 2049

QY 2154 aaaggaag-ccctcaccctctgttatagagccttaatccctctgtctttaaag 2212

Db 2050 AAGAGAGACCTCCATCCTCTGTAATTAACGCCCTTAATGCCCTGTCTTTACTGTAG 2109
Qy 2213 ttaact-agaatcatttttgaagcagcgtgtagagtcctgtaaatgaatgctcgtgac 2271
Db 2110 TTACTTAGATCATTTTGGAGGAGCGCTGTAGAGTCTCTTAATGATGCTCTGGGC 2169
Qy 2272 tagatagagcttgggaacctgctgcttctgtagaagaaacttggccttggcagcag 2331
Db 2170 TAGATACAGCTTGGAGAACCTCTGTGGCTTTTGAACACACTTGGGCTTGGCAGCAG 2229
Qy 2332 caagaggtgaagcgaagcgaacttaactccttcccttccctccctccctccctccct 2391
Db 2230 CAAGAGGTGAAGCGAAGCCACTTACTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCGTA 2289
Qy 2392 ggcacccgaacttggagagacccgctgctgctgtaataactcactcctcctccctccaa 2451
Db 2290 GGCACCCGACACTTGGAGAGACCCGCTGTGTAGTACTTCCATCTCTCTTCCCAA 2349
Qy 2452 gacagatcccaaggaacttaactccttctgctgctgctgctgctgctgctgctgctg 2511
Db 2350 GAGCAGACCCCAAGGCAATTACTCTCTGCTGTCTGCTGCTTATCTGCTGCTGCTGCTG 2409
Qy 2512 cgttagagagcctccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2571
Db 2410 CGCTAGAGACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2469
Qy 2572 agccctcctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2631
Db 2470 AGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2529
Qy 2632 atggagggcagctgaggggaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2690
Db 2530 ATGGAGGGCCACCTAGGGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2589
Qy 2691 -gacacccgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2749
Db 2590 GACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2649
Qy 2750 tggatgatctaatgttagaattcctaattgttagaattcctccttaataatgaatga 2809
Db 2650 TGGATGATTTCTAATGTTAGAAATCTTAATGTTAGAAATCTTCTTAATGATGA 2709
Qy 2810 ttcgtcttctctataattctactactatggtggtgctgtgtgtgtgtgtgtgtgtgt 2869
Db 2710 TTCTCTTCTCTAATCTAATCTAATCTAATCTAATCTAATCTAATCTAATCTAATCTA 2769
Qy 2870 acaacatttaccctcctccttcaactagagaataaacttggtttagaataaaaaa 2929
Db 2770 ACAACCAATTTACCTCTCTTCAAACTAGAGAAATTAAGATTGGTTTGAACCTGAGAA 2829
Qy 2930 aaaaaaaaaaaaaaaaaaaaaa 2954
Db 2830 AAAAAAAAAAAAAAAAAAAAAA 2854

RESULT 4
LOCUS HSA245553 2878 bp mRNA PRI 18-AUG-1999
DEFINITION Homo sapiens mRNA for AP4 zinc finger protein.
ACCESSION AJ245553
VERSION AJ245553.1 GI:5748564
KEYWORDS AP4 gene; AP4 protein; zinc finger protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2878)
AUTHORS Dobner,T.G., Fischer,M. and Grothl,P.
TITLE Cloning of a novel zinc finger protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2878)
AUTHORS Dobner,T.G.
TITLE Direct Submission

JOURNAL Submitted (17-AUG-1999) Dobner T.G., Institut fuer Medizinische
Mikrobiologie und Hygiene, Universitaet Regensburg,
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, GERMANY
FEATURES
source
1. 2878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Hela"
gene
106..1776
/gene="AP4"
106..1776
/note="zinc finger protein"
/product="AP4 protein"
/protein_id="CA53100.1"
/db_xref="GI:5748565"
CDS
BASE COUNT 528 a 973 c 839 g 538 t
ORIGIN
Query Match 89.9%; Score 2654.6; DB 77; Length 2878;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 2772; Conservative 0; Mismatches 49; Indels 8; Gaps 8;
Qy 107 catlctcagagagaagaacgatgctgtaagcgtctgtcaggggccccctggcattgg 166
Db 51 CACCTCAGCAGAGAGAAACCGATGCTGGAACGTCGTGCA-GGGCCCTGGCCATGG 109
Qy 167 gcttgcccgagcccgagcctcttctgagccctcccaaggaatgcaacccagaccctggga 226
Db 110 GCCCGCCAGCCCGGACCTCTTCTGCGCTCCAGAGATCACCCAGACCTGGGGA 169
Qy 227 aggaatcccgagctgagcaagaagcagctcagtgccagctgctggtgccagccc 286
Db 170 AGGAGTCCCGGGCGGTAGAGCAACAAGCAGTCAGTGCCAGTGTGCCAAGCCC 229
Qy 287 caggcagggccatcgtctgccaactgtcgaaggcaactccctggtggtgtgtgtgt 346
Db 230 CAGGACAGGCCCATGCTGTGCTGCAAGCAGTCCCTGCTGGGTGGCTCTGT 289
Qy 347 ggtctcaaccccgagctgagccagctgctgctgctgctgctgctgctgctgctgct 406
Db 290 GGGCTTACACCCGGCGGTGCGAGGCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 349
Qy 407 gcttcgcacatgcccctctttagactgacagcagcaggtccatgctgctgcccaccag 466
Db 350 GCTTTCGCATGCCCTTCTTAGCAGCTGACGGCAGGTTCATGCTGTGCAACCCAG 409
Qy 467 acctgggcttggcctgcaactctgtggtgagagcttccagagctggtgtgctgtgtc 526
Db 410 ACCTGGGCTTTCCTGCACTCTGTGGGAGAGCTTCCAGAGCTGGGTGGCTGCTGCTTC 469
Qy 527 tgcactgctgagccattcagctgtaagaagcaaccatgctgtgtccaaatgagagaac 586
Db 470 TGCATCTGGGGCCATTTCAGTGAAGGGCCCATGCTTGTCCAAATGCGAGAAC 529
Qy 587 gcttcgcaagaagaagctgagctcactcctcagcgtgtgtgcaacctcccgccgg 646
Db 530 GCTTTCGGAGAGAGAGAGCTTCAAGCTCATCTGACGGGTGCCACCTCCGGCCGG 589
Qy 647 agcccgagccttcatatgagcaactgtgagcagagcttggccagtggagcagctag 706

QY	2860	actaaacgaacacacattaccctccctttcaactagaaataacatttgttta	2919
Db	2810	ACTAAACAGAACACCACTTACCCTCCCTTTCAAACTGAGAAATTAAGATTGGGTTT	2869
QY	2920	gaaataaaaa	2928
Db	2870	AGAAAAAAA	2878
RESULT	5		
AC073111			
LOCUS			
DEFINITION	AC073111 179225 bp DNA	HTG	15-JUL-2000
	Homo sapiens chromosome 7 clone RP11-511P7, WORKING DRAFT SEQUENCE		
VERSION	31 unordered pieces.		
AC073111			
AC073111.3	GI:9211463		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 179225)		
JOURNAL	Waterston, R.H.		
REFERENCE	The sequence of Homo sapiens clone		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 179225)		
JOURNAL	Waterston, R.H.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (08-JUN-2000) Genome Sequencing Center, Washington		
TITLE	Mo 63108, USA		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis		
COMMENT	On Jul 15, 2000 this sequence version replaced gi:8572520.		

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0511P07
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160317 bases at least Q40
Consensus quality: 166937 bases at least Q30
Consensus quality: 169751 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 176225; sum-of-contigs
Quality coverage: 4.24 in Q20 bases; agarose-fp
Quality coverage: 4.40 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1315: contig of 1315 bp in length
*      1316      1415: gap of unknown length
*      1416      2954: contig of 1539 bp in length
*      2955
*      3055      3054: gap of unknown length
*      4541: contig of 1487 bp in length
*      4542      4641: gap of unknown length
*      4642      6249: contig of 1608 bp in length
*      6250      6349: gap of unknown length
*      6350      8039: contig of 1690 bp in length
*      8040      8139: gap of unknown length
*      8140      9473: contig of 1334 bp in length
*      9474      9573: gap of unknown length

```

*	9574	11903: contig of 2330 bp in length
*	11904	12003: gap of unknown length
*	12004	14193: contig of 2190 bp in length
*	14194	14293: gap of unknown length
*	14294	17408: contig of 3115 bp in length
*	17409	17508: gap of unknown length
*	17509	20169: contig of 2661 bp in length
*	20170	20269: gap of unknown length
*	20270	22745: contig of 2476 bp in length
*	22746	22845: gap of unknown length
*	22846	26671: contig of 3826 bp in length
*	26672	26771: gap of unknown length
*	26772	30438: contig of 3667 bp in length
*	30439	30538: gap of unknown length
*	30539	33209: contig of 2671 bp in length
*	33210	33309: gap of unknown length
*	33310	36685: contig of 3576 bp in length
*	36686	36685: gap of unknown length
*	36986	41449: contig of 4464 bp in length
*	41450	41549: gap of unknown length
*	41550	45860: contig of 4311 bp in length
*	45861	45960: gap of unknown length
*	45961	49359: contig of 3399 bp in length
*	49360	49459: gap of unknown length
*	49460	54617: contig of 5158 bp in length
*	54618	54717: gap of unknown length
*	54718	61279: contig of 6562 bp in length
*	61280	61379: gap of unknown length
*	61380	66606: contig of 5227 bp in length
*	66607	66706: gap of unknown length
*	66707	73439: contig of 6733 bp in length
*	73440	73539: gap of unknown length
*	73540	80804: contig of 7265 bp in length
*	80805	80904: gap of unknown length
*	80905	87301: contig of 6397 bp in length
*	87302	87401: gap of unknown length
*	87402	95200: contig of 7799 bp in length
*	95201	95300: gap of unknown length
*	95301	106211: contig of 10911 bp in length
*	106212	106311: gap of unknown length
*	106312	114409: contig of 8098 bp in length
*	114410	114509: gap of unknown length
*	114510	124101: contig of 9592 bp in length
*	124102	124201: gap of unknown length
*	124202	136041: contig of 11840 bp in length
*	136042	136141: gap of unknown length
*	136142	155777: contig of 19366 bp in length
*	155778	155877: gap of unknown length
*	155878	179225: contig of 23348 bp in length

FEATURES

SOURCE

1. .179225

Location/Qualifiers

```
misc_feature      1. 1315
/note="assembly_name:Contig25
misc_feature      1416. 2954
/note="assembly_name:Contig27"
misc_feature      3055. 4541
/note="assembly_name:Contig28
4664. 6249
/note="assembly_name:Contig29"
6350. 8039
/note="assembly_name:Contig30"
8140. 9473
/note="assembly_name:Contig31"
9574. 11993
/note="assembly_name:Contig32"
12004. 14193
/note="assembly_name:Contig33"
14244. 117408
/note="assembly_name:Contig34"
```

```
misc_feature      17509. .20169  
/note="assembly_name:Contig35"  
misc_feature     20270. .22745  
/note="assembly_name:Contig36"  
misc_feature     22846. .26671  
/note="assembly_name:Contig37"  
misc_feature     26772. .30438  
/note="assembly_name:Contig38"  
misc_feature     30539. .33209  
/note="assembly_name:Contig39  
clone_end:r7  
vector_side:right"  
misc_feature     33310. .36885  
/note="assembly_name:Contig40"  
misc_feature     36986. .41449  
/note="assembly_name:Contig41"  
misc_feature     41350. .45860  
/note="assembly_name:Contig42"  
misc_feature     45961. .49359  
/note="assembly_name:Contig43"  
misc_feature     49460. .54617  
/note="assembly_name:Contig44"  
misc_feature     54718. .61279  
/note="assembly_name:Contig45"  
misc_feature     61380. .66606  
/note="assembly_name:Contig46"  
misc_feature     66707. .73439  
/note="assembly_name:Contig47"  
misc_feature     73540. .80804  
/note="assembly_name:Contig48"  
misc_feature     80905. .87301  
/note="assembly_name:Contig49"  
misc_feature     87402. .95200  
/note="assembly_name:Contig50"  
misc_feature     95301. .106211  
/note="assembly_name:Contig51  
clone_end:SP6  
vector_side:left"  
misc_feature     106312. .114409  
/note="assembly_name:Contig52"  
misc_feature     114510. .124101  
/note="assembly_name:Contig53"  
misc_feature     124202. .136041  
/note="assembly_name:Contig54"  
misc_feature     136142. .155777  
/note="assembly_name:Contig55"  
misc_feature     155878. .179225  
/note="assembly_name:Contig56"  
CASE COUNT      47914 a 40487 c 42219 g 45578 t 3027 others  
ORIGIN
```

Query Match	42.5%	Score 1256.4	DB: 58	Length 179225
Best Local Similarity	98.2%	Pred. No. 3.7e-185		
Matches 1344	Conservative	0	Mismatches 16	Indels 8
				Gaps 7
QY 1561	cggcgatcatcacacacggcgagagaagcccttaagctctgcccgaactgcygcaaaagccttaagc	1620		
Db 124221	CTGCGCATTCACACACGGCGAAGAAATCCCTACGTCGCCGCACT--TGCACAACTTCACG	124278		
QY 1621	cagaagttcaacactgtgttgcacacggcgatcatcacaaagcgagcgagcgcccttaagctgt	1680		
Db 124279	CAGAAGTTCACACTGTGTGTGTCGACCGCGGCATTCACACGGGCGAGCGGCGCTTACGCTTCT	124358		
QY 1681	ccc_gacttcgacacggcgaacttaagccagaagtctcaacactcatcacccacgcgaagaagcca	1739		
Db 124339	CCCNAGTCGCGACCGACGACTTCAGCCAGAGTCCAAACCTATCACCCACCGCAAGAGCCA	124398		
QY 1740	catcgggagacggcgacctctgcttctgccaactgtgtggcgaagacttcgaagaagaagaag	1799		
Db 124339	CATCGGGAGACGGCGCCCTTCTGCTGTGTGCCATCTGTGTGCCAGACCTTCGACGACGAGGAGAG	124458		
QY 1800	actcctggcccacacagaagaacgcatgtctlgaagaagtlggcgggcgccgtgttgcgtg	1859		

Dh	124459	ACTCTGGCCCAACGAAAGAAAGCACGATGTCGTGAGACGGTGGCGGGCCGTGGCTG	124518
Oy	1860	agagaaggcttgaggcttccttcgtgttgaggatcgaagtcgagcttgagggttgcttcctagtg	1919
Dh	124519	AGAGAGGGCTGGGGGTCCTGCTGGTGGAGATCGGACATGGGGCTGGGGGTGCTGCTTAATG	124578
Oy	1920	ctggaataaggaggacaatggaaatccctatagagggaatggaaatgcgagggaatgagctgggt	1979
Dh	124579	CTGGATGAGGGGACAAATGGGAATCTCTAGAGGGATGTGMAAACCGGGGAGATGAGCTGGT	124638
Oy	1980	ggggccctctagcgaaggaaggttaaccgggttgccaggggaaccccttcgaagcgcaag	2039
Dh	124639	GGGGCTCTTAGCGAGAGAGGTCAACCCCGGTGGCCAGGGAACCCACTTCAAGCGGAG	124698
Oy	2040	gacgcggagccctccagagctggtgtgtgtataagctccgtccctgactgacctgtgacctggaa	2099
Dh	124699	GACGGCGGCTTCAGATGAGTGTGTGTGCTAAAGGCTCCGTCCTGACTGCCCTTGCCCTGGAA	124758
Oy	2100	aagcagaata-catccgccccctacagccctctgctatagaggacacagctgaaag	2158
Dh	124759	AAGCAGCAATAGCATCCGCCCTTAGAGCCCTCGCTGCTAAGAGACCCACAGTGAAGG	124818
Oy	2159	aag-cccccatctctctgtgataaagccttaagccccctgtctttacatgaattact	2217
Dh	124819	AAGACCCCTCAATCCCTGTGTATTAACCCCTTAAGCCCGTGTCTTAACTGTAAATTA	124878
Oy	2218	t-agatcaatcttgaaagcaaggctgtgtatagcttcgttaatatgaatgccttgaggatag	2276
Dh	124879	TAAAGTCAATTTTGGAGAACAGCGCTGTGTAAGTCTCTTAATATGAATGCTCTGGGCTA	124938
Oy	2277	acagcttgagaacctgcgtgacctgtgtatagacagaacttggaccttgcacagcagaaga	2336
Dh	124939	ACAGCTTGAGAAACCTGCTGGCTCTGTGTAAGACAGCACTTGGGGCTTTGCCAGACGA	124998
Oy	2337	ggtgaagcgaaagccaactcttaacctctccctccctccacatgcgccccctgtgtaaggcac	2396
Dh	124999	GGTGAAGGAAAGCCACATCTTAACCTCCCTCCCTCCACATGGCCCGCCGCTAGGCA	125058
Oy	2397	ccaagcttgaggaagaccgctgtctgttaataacttcaactccactctctcccaagaagca	2456
Dh	125059	CCAGACTTGGAGAACCCGCTGTGCTGTAATACTTCCATCTCTCTCTCCCAAAGACA	125118
Oy	2457	gatacccaaggcatctactcctctgtctgcctgccttatctgtctgcgccccctccagcgctg	2516
Dh	125119	GATCCCAAGGCAATTAATCTCTGTGGTGTGTGTCGCTTATATCTGTGGGCCCTGCCAGGGCT	125178
Oy	2517	agaagctccctgcgtctcagagaacatgtgtccaggctctgtctgaacacgcgacgcc	2576
Dh	125179	AGAGCTCCCTCGCTGCTACACACACATGTGTCCAGGCTCTGTGTGMAACCGCAGGCC	125238
Oy	2577	ctccctgcctctccccaagcttaagcatgttcaacgggaagagctgcgcgcatctgtgtatg	2636
Dh	125239	CTCTCTCCCTCTTCCAGAGCTCAACATGTCAACGGCAAGGATCTCCCATTTGGTGGGA	125298
Oy	2637	gggccaagcttgagggaagtgtgctgtgagtcttc-ttcttccatcttctaagcatat-gaca	2694
Dh	125299	GGGCGACGCTGAGGGGAAGTGTCTGTGTAGTTCCTTTTGTCCATTTCTACCATATGGACA	125358
Oy	2695	ccctgagctctgacttgaagcaattaggtgaacagaaactccgaacctccctgagcccttgat	2754
Dh	125359	CCTGGCCCTCTGCTGAGCACTTAAGTGAACAGGAATTCGCCACCTCTCGAGGCCCTGAT	125418
Oy	2755	gatttcaatgtttgaaatcttaattgttgaaatcttctctaataatgaatgaattctg	2814
Dh	125419	GATTTAATTTTGTGAATAATTTTAATTTGTGTGAATAATCTTCTTAATTAATGAATAATTCG	125478
Oy	2815	ctttcctaataatcttaactaatttgagcgtgtctgtctctctggaactaaacgaacaac	2874
Dh	125479	CTTTCTTAATTTTCTACACTTTGTGGGCTGTGTGTGTCTGTGAACCTAAACGAAACAC	125538
Oy	2875	catttaacctctcttccaactagagataaacaattggttttagaa 2922	

QY	1947	ttacacaaataagcccatcttgacttcgcacccggcgcatccacaacccgggagaaagccctaac	1006
Db	132665	TCACCAAAAGCCCTTACTGTACCTGCACCCGGCCATTCATACCGCGAGAACCCCTTACC	132724
QY	1007	cgtgcagaagatgtcggcccgccgtcttcggcacaacaaccacaactctgttctcaecagcaaga	1066
Db	132725	CATGCACCGAGTGTGGCCGGCTTCGGCCACAAACCCAACTCTGTGTGCGCAGGCATAAA	132789
QY	1067	ttcaaaagcgatcgcaggaggtctcgcccagcgcccccgcggcgaggagcccccaactctgc	1126
Db	132785	TCACAAACCGCTTGGAGGTCTTCACAGCGCAGGCTGCCACACACCCGGAAGTCAACAGATTG	132844
QY	1127	cagcgcgcccccaagagatcgcgcgcgcgacccaccccgcgcgtatacctctgaaccggccc	1186
Db	132845	CAGCAGACCTTATG-----132855	
QY	1187	aggagccgcgcgcgaaggccccgcgcagagaccccgacagaccgatgcgaagccccccct	1246
Db	132860	CACAACCTGTGACTTGGGTGGCCCTMGGGGTCTCCCGCGGACCCAGCTGAGGCACTCTGCG	132919
QY	1247	ccctctaaagcttcgcagcagactctgggcaggagacttcgcgctctgagagcttccctcggccc	1306
Db	132920	TCCTTCATATGCTGTCCACACGCGCGCACCTTCCGGCTGGAACGCTCTCGCGCTTAC	132979
QY	1307	accagcggcacaagacccggggagcgcccccttcaactctgcgcgcagatltggggaaagaaacttg	1366
Db	132980	ACCAAGGCGACACACAGAGGGAGAGGCCCTTCTGCTGCACAGATGTGGCAAGACATTGG	133039
QY	1367	gcagaagaacgcacacttgtgtgcgaacttcgcgcgtgcactccgcggagacgcgccttcgcct	1426
Db	133040	GCAAGAAAGCACACTGTGTGGCGCACTCACCGGTGCACTCCGGGAAAGTCCCTTGCCT	133099
QY	1427	gcgagagatgcgcgcgcgcgtcttcccaaggcagacatctgycgcgcatcgcgcgcgagacc	1486
Db	133100	GGAGAGATGTGTGTGCGCGCTTTCACAGGGCACCACTGTGGCACCCACCGGAGAACCC	133159
QY	1487	acgcccccgatccgcgcctctgtgtgtcccgactctgcgcgaagccttcgcgcacaacaacct	1546
Db	133160	ATGCACACGAGAGGCCCTTCTGTGTGCGGACTCGGCACTGGCAGAGCTTTCGGCCACAGCCCT	133219
QY	1547	accttgcgcgcgcacccgcgcacatccacacccgcgcgcgaagcccttaactgtctgcgccgactgc	1606
Db	133220	ACCTTGCCTGCGCACCGACGACCTTCACACAGGCGAAGAACCTTATGTGTGTCGCCACTGTG	133279
QY	1607	gcgaagccttaagccagaagtccaaacttgtgtgcacccggcgcatccacaacggcgagc	1666
Db	133280	GCAAAAGCTTTCAGTTCAGAAATCCAAACCTGGTGTCCCAACCGCGCATCACACAGGCGAGC	133339
QY	1667	ggcccttaagctgtcccgactctgcgcacgcgcacatcccaacgaagaatccaaactcaacc	1726
Db	133340	GGCCCTTACGCCTGCCCGCACTGTGTATCTGTACTGTCAAGATCCAAATCTTATTCACAC	133399
QY	1727	accgcgaagacccacatccgcgcgcgcgcctctctgtctgtccatctgtgcgcagaactctgc	1786
Db	133400	ACCGGAAGAGCCACATCCGGATGGCGGCTTCTGTGTGGCATGTGTGGCAAGACCTTTG	133459
QY	1787	acgcacgagagagagactctctgtgcccaacgaagaagacagatgtctgaagacgltgycggg	1846
Db	133460	ATGAGGAGAGACCGACTCTTGTATGACACGAAAGAGCATGTAGCTGTGAGAGGGGCGCTTAGG	133519
QY	1847	ggccggttgcgttgagagagggccttgggtctcctctgtgttgagagatcgcgaagtgcgccttgagg	1906
Db	133520	CCTAGGAAGACAGCCCTGGAATATCTTTGGCAACAGGGGTGTCAAGTGG--TCCGGGTGGGGT	133577
QY	1907	tgacctgcctagtgtctgtagatagggagacaatltggaactcctaagaggggatatggaagatgcggg	1966
Db	133578	ACCTTCACTGTATGCTGTAGAGAAACAGGTAAAGATCCCAAGAGGATAGAGGAGACAGTG	133637
QY	1967	gagtgagcttgggttggccctctctgagagagaggtcaacccccgttgcgcagggaacccac	2026
Db	133638	ATCTGGCTCTGT-----TTAGGAATGACATTTCCCTTAGGTATCCAGGGAAACCAT	133689

Db 1428 CAACACAGGAGCATCCACCGGCGAGCCCTACATCTGCGCGAGCTGCGCAAG 1487
1530 cttccgacacaacccctcctcggcgacggcgatccacacggcgagaaacctta 1589
Db 1488 CTTATCATGTGAGCTCCACCTTATCTCCGACGCAATCCACACCGGTGAGACCCCTTA 1547
1590 cgtctgcccagctcggcacaagcctcagcacaagtcacaacctgctgctcagcg 1649
Db 1548 CAAGTGTCCAGCTGCGGCAAGAGCTTATCCGACGCTCCACCTTATCCAGACCGCG 1607
1650 catcacacggcgagcgagcgacctacgctgtccgactcgacggcgagcttcagcaga 1709
Db 1608 CACGCACACCGCGGAGAGCCCTACAGTCCCGCCGAGTGGCGGAGAGCTTCAGCGAG 1667
1710 gtccaaacctatcccccacggcagaagaccatccggagcgcgctctcgtgcat 1769
Db 1668 CTCACACCTTATTTACCCGACGCTCCGACGCACTGAGAGAACTGTCTGCTCGGA 1727
1770 ctgtgagcagacacctcgcagcagagagactcctgagcccgacagaaagcagca 1826
1728 CTGGGGAGAGGCTTCCCTGGAGAGCCAGAGCTGAGAGCAGCAGCGGTGATCATGA 1784

RESULT 9
AC013570/c
LOCUS
DEFINITION Homo sapiens clone Rp11-2C24, WORKING DRAFT SEQUENCE, 37 unordered
pieces.
AC013570
VERSION AC013570.4 GI:8072509
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 185306)
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Bara, N., Beckler, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dekrellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
Ferreira, P., Fitzhugh, K., Forrest, C., Funke, R., Gage, D., Hooton, L.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karalis, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McDonald, J., Meidrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced g1:7654731.
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2523
Center clone name: 2_C_24
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
Consensus quality: 162201 bases at least Q40
Consensus quality: 173731 bases at least Q30
Consensus quality: 178289 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 181706; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1008: contig of 1008 bp in length
* 1009 1108: gap of 100 bp
* 1109 2471: contig of 1363 bp in length
* 2472 2571: gap of 100 bp
* 2572 4482: contig of 1911 bp in length
* 4483 4582: gap of 100 bp
* 4583 6444: contig of 1862 bp in length
* 6445 6544: gap of 100 bp
* 6545 8758: contig of 2214 bp in length
* 8759 8858: gap of 100 bp
* 8859 10776: contig of 1918 bp in length
* 10777 10876: gap of 100 bp
* 10877 12177: contig of 1301 bp in length
* 12178 12277: gap of 100 bp
* 12278 13361: contig of 1084 bp in length
* 13362 13461: gap of 100 bp
* 13462 15267: contig of 1806 bp in length
* 15268 15367: gap of 100 bp
* 15368 16785: contig of 1418 bp in length
* 16786 16885: gap of 100 bp
* 16886 18775: contig of 1890 bp in length
* 18776 18875: gap of 100 bp
* 18876 21003: contig of 2128 bp in length
* 21004 21103: gap of 100 bp
* 21104 22763: contig of 1660 bp in length
* 22764 22863: gap of 100 bp
* 22864 25665: contig of 2802 bp in length
* 25666 25765: gap of 100 bp
* 25766 28291: contig of 2526 bp in length
* 28292 28391: gap of 100 bp
* 28392 29739: contig of 1348 bp in length
* 29740 29839: gap of 100 bp
* 29840 33467: contig of 3628 bp in length
* 33468 33567: gap of 100 bp
* 33568 37139: contig of 3872 bp in length
* 37140 37239: gap of 100 bp
* 37240 41052: contig of 3813 bp in length
* 41053 41152: gap of 100 bp
* 41153 45504: contig of 4452 bp in length
* 45505 45604: gap of 100 bp
* 45605 50022: contig of 4418 bp in length
* 50023 50122: gap of 100 bp
* 50123 54849: contig of 4727 bp in length
* 54850 54949: gap of 100 bp
* 54950 59045: contig of 4096 bp in length
* 59046 59145: gap of 100 bp
* 59146 65075: contig of 5530 bp in length
* 65076 65175: gap of 100 bp
* 65176 71663: contig of 6488 bp in length
* 71664 77852: gap of 100 bp
* 77853 77952: contig of 6089 bp in length
* 77953 83219: gap of 100 bp
* 83220 83319: contig of 5267 bp in length
* 83320 88571: gap of 100 bp
* 88572 88671: contig of 5522 bp in length
* 88672 88671: gap of 100 bp


```

source
/chromosome="16"
/map="p13.3"
/clone="YAC 26fe7"
/sub_clone="30c3"
90616. .137018
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="p13.3"
/clone="YAC 26fe7"
/sub_clone="30e1"
126806. .170318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="p13.3"
/clone="YAC 26fe7"
/sub_clone="30g4"
complement(161861. .161970)
/standard_name="CHLC.ATRA1E08.P34246"
163476. .164414
/gene="hsol1mf"
163476. .164414
/gene="hsol1mf"
/codon_start=1
/number=1
/product="olfactory receptor"
/protein_id="CA05903.1"
/db_xref="GI:2808658"
/db_xref="SPTREMBL:O43749"
/translation="MSGTQSSVSEFLLGLSRPQOHLFEFLSMYATVIGNLL
ILTASIDISCLHTPMYEFELSNLSEVDICFETTPKMLAHILLETQTSRCGCLTOY
EFMEFVADNDELTAAMADHFAVACHPLHTAKMTHQCALVGLWVNLNLITLT
ELMAPLSCADNALTIFPCDVTPLPLKISCDTHNEVTLISEGLVNMTPICLIASY
MHTQTVKAVSTGKGRKAFSTGCSGLAVLLFYSTIIAYFNPLSSHSAEKQMTNAV
LYTQVTPMLNPFYLSLRNRYLKGALKKRVGKVESV"
173301. .217867
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="p13.3"
/clone="YAC 26fe7"
/sub_clone="30f9"
174780. .175778
/gene="olfm2"
174780. .175778
/gene="olfm2"
/codon_start=1
/pseudo
/number=1
/product="olfactory receptor pseudogene"
/db_xref="PID:e1310000"
179596. .222837
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="p13.3"
/clone="YAC 26fe7"
/sub_clone="30e10"
complement(183121. .194589)
/gene="zn1mf"
complement(join(183121. .183842,191650. .191773,
192062. .192150,192735. .192984))
/gene="zn1mf"
/codon_start=1
/product="zinc finger protein"
/protein_id="CA05905.1"
/db_xref="GI:2808659"
/db_xref="SPTREMBL:O15361"
/translation="MMAKVPMPPKPSFLIYPPDSKIGODLDATNGPTTHO
LVLEHFLIFLPRPSLVQSKYKETLVIMKRVSSSLQNRVPRPLVKILPGVQKEDE
VLSLVKANPELVNEDLVNFHCEVSLDPTQQLTSEKEDSSVSGMILVANGSN
PGVEDEPERPEVNEDEKSSDDDEMDSILVSQPPDQEKERLNTSIPQKKRRNLL

```

[illegible]

Df	48598	GCAGCCACACGGGCGAGCGGCCCTTACAGGTGTCAGACGTCGGGTGCACCTTCGCGCGCA	48539
Oy	1793	aggagaagactctgtgccaccacgaagaagacg	1825
Dd	48538	ACGCCATCTGCGGGCGCACCTGGCTACCANG	48506

RESULT 11

AKO23017

LOCUS AKO23017 2382 bp mRNA PRI 29-SEP-2000

DEFINITION Homo sapiens CDNA FLJ12955 fis, clone NT2RP2005496, moderately similar to ZINC FINGER PROTEIN 135.

ACCESSION AKO23017

KEYWORDS AKO23017.1 GI:10434737

SOURCE Oligo cloning; fls (full insert sequence).

ORGANISM Homo sapiens

REFERENCE Homo sapiens teratocarcinoma cell_line:NT2 CDNA to mRNA, clone_11b:NT2RP2 clone:NT2RP2005496.

AUTHORS Homo sapiens

TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (sites)

JOURNAL Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K., Tanai,H., Kinata,M., Matanabe,M., Hiraoka,S., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuko,Y. and Kanehorri,K.

AUTHORS NEDO human cDNA sequencing project unpublished (2000) 2 (bases 1 to 2382)

TITLE Isogai,T. and Otsuki,T.

JOURNAL Direct Submission Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

COMMENT NEDO human cDNA-sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES Location/Qualifiers

SOURCE 1..2382 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="NT2" /cell_type="teratocarcinoma" /clone="NT2RP2005496" /clone_11b="NT2RP2" /note="cloning vector: PME18SFL3-mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."

BASE COUNT 487 a 789 c 656 g 450 t

ORIGIN

```

Query Match          10.7%;   Score 315.2; DB 37; Length 2382;
Best Local Similarity 56.5%; Pred. No.1.8e-39;
Matches 606; Conservative 0; Mismatches 463; Indels 3; Gaps

```

756 caaagcctctgggcccccgcccaaggcgccgcgcgcgtgaccccccgcgcgcgcgttg 815
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db CAGCGCGCGCTGGGGGAGCGCCGAAAGAGAGCATCCTGCACGCCCGCAGGAGAGAG 147

816 agatgcgctgcacgcgcaccttcagtgtgctgttgttgcaagcgcttcgcgcacaagc 875
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db TGGAGCGCCCCAAGAAGCCGTGGGAATGCGGGAGCTGCGGGAAGCGCCTTGACTGTTGC 207

876 caacttgtatgcctacgcgcgcgtgcaaacaggcgagcgcccccaacagttgcgcagatg 935
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db CGCGTTCAATCTTACACCGAGAGATCCACACCGGGGAGAGAACATTGTCGTCGCCGAGTG 267

[illegible]

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Sakalidis,G., Burkhart,Schultz,K., Gordon,L., Dias,J., Scott,D., Stillwagen,S., Phan,H., Velasco,R., Do,L., Regala,M., Terry,A., Dangnan,L., Erler,A., Christensen,M., Georges,A., Avila,J., Atlix,C., Andreise,T., Antco-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,K., Thomas,P., Quan,G., Krommiller,B., Ariellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carraro,A.V.
TITLE	Sequence analysis of a 3.6 Mb region in 19q13.4 between D19S891 to ZNF134
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 130067)
AUTHORS	Lamerdin,J.E.
TITLE	Direct Submission
JOURNAL	Submitted (06-APR-1999) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT	Map and sequence oriented from q centromere to telomere. BAC 37295 (CIT-B-21A4) overlaps cosmid R31665 (AC005498) to the left from bases 1 to 10,159 of this accession, and is separated from BAC 330812 (CIT-B-470F8; AC006115) to the right by a sequence gap of unknown size. Additional chromosome 19 map and sequence information are available at: http://www.bio.livnl.gov/bjrrp/genome/genome.html.
FEATURES	Location/Qualifiers
SOURCE	1. 130067 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CIT-B-21A4" /chromosome="19" /map="19q13.4 between D19S891 and ZNF134" /clone_lib="CIT-B Cal Tech BAC library" /note="LIML Clone Name: BC37295" complement(4..320) /rpt_family="MUT1B" join((332..509,682..777,1833..1996,2942..3056,3385..3480,6884..8266) /note="Hypothetical ZNF-like human protein" /codon_start=1 /evidence=not_experimental /product="R31665_2 [AA 1- 673]" /protein_id="AAD23606.1" /db_xref="GI:4567178"
CDS	/translation="MGHEHMLSKNNVTSFOGLVTFGDVAVDQSEEWMLNPIDRL YRKVMLENYRNLSIGLCVSPDVISLSEGGKEPWTYVKMTWPCPLKAWMKIKEL PLKKDCDEGKLISOAVTERLTSTNLEYSLIGHMYDLAEFPQGLVTKNLAVDFRO QLPHPAKNCCKNGIWMNSNDLCSAGCVAKPRVLVILDEKPMWVRRELIGSLFSGI YACKKLFKNECKKTITQSSSLTVHORHTGKPYKCNGCAEFDGSSFKAHOCFT GKRYPBCIECGKAFFQNTSLIRHWRYTHGEKPEFCIDCGAFSDHIDLNOHRHTHG EKREYKCDVCHKSFRYGSJLVHORHTHTGEPCEVCGRKAFSHASLVOHRSVGSKY PFKSCGEGKAFRONIHASHLRHTHTGEPCEVCGRKAFSHASLVOHRSVGSKY ECVGAKAFQOKNAHLAOHQHTHGEKPYECCKGKAFQSOTTHLIOHOHVHGEKPYKC MECCGAFQGNSSCTOQHRLHTGORPYECIECKGKATKSSSLCHRSHRTGERPKESY CGAFSHROSLSVHORISHGKRPYECKECKKFTFIQIHLNOHKRVHTERSTNYKSKR KVRFQHTLAHORHTHTGESSYCPSLPSTSNPDVLEPFKLMWPSLPSPP"
repeat_region	/rpt_family="AluYo" complement(4953..4996) /rpt_family="AT_rich" complement(5673..5793) /rpt_family="(GAAA)n" complement(5799..5866) /rpt_family="LIMB5" 5864..5996 /rpt_family="LIMAG" complement(6099..6414) /rpt_family="AluJo" 8415..8489 /rpt_family="MIR" complement(8977..9293) /rpt_family="MER58B" complement(9302..9497) /rpt_family="MER58A"
repeat_region	complement(9792..10090) /rpt_family="Alusg1" complement(10096..10278) /rpt_family="AluBo" complement(12206..13159) /rpt_family="MIR5" complement(13352..13406) /rpt_family="(GCAG)n" complement(13407..13697) /rpt_family="Aluub" complement(15043..15575) /rpt_family="MUT1f" complement(16011..16143) /rpt_family="LIP4" complement(16194..16338) /rpt_family="MUT1-INTERNAL" complement(16573..16870) /rpt_family="MUT1-INTERNAL" complement(17315..17488) /note="BLASTn similarity to G44255 (13..186); match: 0.99 score: 1.3e-63; database searched: month.na; WtAF-3903-STS Human Tudson SANGER Homo sapiens STS genomic, sequence tagged site [Homo sapiens]" complement(17491..17766) /rpt_family="AluBo" complement(17772..18156) /rpt_family="LIM4" complement(18162..18925) /rpt_family="LIM4" complement(19178..19553) /rpt_family="MUT1A2" 19599..19622 /rpt_family="AT_rich" complement(19623..19726) /rpt_family="LINE2" complement(20008..20389) /note="predicted exon, program: graill2exons_human_1.3, frame: 2, quality: good, score: 57.000" complement(20721..20768) /note="predicted exon, program: graill2exons_human_1.3, frame: 0, quality: excellent, score: 100.000" complement(21509..21720) /rpt_family="MIR" complement(21729..21817) /rpt_family="MIR" 21897..21996 /rpt_family="MIR" 22540..22611 /rpt_family="AluBo" 22671..22874 /rpt_family="AluJo" join((23186..23245,27268..27607,30057..31456) /note="Hypothetical ZNF-like human protein" /codon_start=1 /evidence=not_experimental /product="BC37295_1" /protein_id="AAD23607.1" /db_xref="GI:4567179"
repeat_region	/translation="MKSDEVEVAGIKICLKANSGLSVTFPTVAIDFODEWMLNLAQ LSIYKRWLENTRYNVSGKSFLLPHDSVFYAIIIRISHSPPOHROPYLVELLAG RSICSPDIVSILEOEKPWVIRKGMNRGICDEKPECEIECGKAFSGSAHLAOHORHT HTGEPCEVCGRKAFSONHLYOHORVHTGEPYCCKGKAFSQLAHLOHOHVHGEK PYECIECKGKAFSDSSLAHHRIHTGTGKPECTIDCGKARONASLIIRHRRYYNGEE KRPDCTIDCGKAFSDHGLIOHKTRHTTGERPPYKANCVCAGFSDGSLTYVHORHTGEP YECNTEKRAFSHRSGJTLHORVHTGEPYECKECKGAROSTHLAHORHTGEPYE CKECSYTESQNAHLAOHORHTHTGEPYECKECKGARFSDIAHYOHORVHTGEPYEC IECKGAFSDSYLVOHORLHSGKRPYECKECKGARFORASLIICHORHTGEPYCNVC GKAFSHRSLSLTIHORHTHTGEPYECKECKGAFSGVAHLTLHKRIHTGTGEPYECKECK AFROSVTLAHORHTHTGESSYTLSSALPYHQVL"
repeat_region	complement(24008..24243) /rpt_family="LIME" complement(24295..24317) /rpt_family="AT_rich"

[illegible][illegible]

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1529: contig of 1529 bp in length
* 1530 1629: gap of unknown length
* 1630 2945: contig of 1316 bp in length
* 2946 3045: gap of unknown length
* 3046 4365: contig of 1320 bp in length
* 4366 4465: gap of unknown length
* 4466 6057: contig of 1592 bp in length
* 6058 6157: gap of unknown length
* 6158 7646: contig of 1489 bp in length
* 7647 7746: gap of unknown length
* 7747 8848: contig of 1102 bp in length
* 8849 9963: contig of 1015 bp in length
* 9964 10063: gap of unknown length
* 10064 11257: contig of 1194 bp in length
* 11258 11357: gap of unknown length
* 11358 12937: contig of 1580 bp in length
* 12938 13037: gap of unknown length
* 13038 14304: contig of 1267 bp in length
* 14305 14404: gap of unknown length
* 14405 15589: contig of 1185 bp in length
* 15590 15689: gap of unknown length
* 15690 16899: contig of 1210 bp in length
* 16900 16999: gap of unknown length
* 17000 18157: contig of 1158 bp in length
* 18158 18257: gap of unknown length
* 18258 19628: contig of 1371 bp in length
* 19629 19728: gap of unknown length
* 19729 21140: contig of 1412 bp in length
* 21141 22406: contig of 1166 bp in length
* 22407 22506: gap of unknown length
* 22507 23647: contig of 1141 bp in length
* 23648 23747: gap of unknown length
* 23748 25186: contig of 1439 bp in length
* 25187 25286: gap of unknown length
* 25287 26291: contig of 1005 bp in length
* 26292 26391: gap of unknown length
* 26392 28358: contig of 1967 bp in length
* 28359 28458: gap of unknown length
* 28459 30663: contig of 2205 bp in length
* 30664 30763: gap of unknown length
* 30764 32618: contig of 1855 bp in length
* 32619 32718: gap of unknown length
* 32719 34502: contig of 1784 bp in length
* 34503 34602: gap of unknown length
* 34603 36285: contig of 1683 bp in length
* 36286 36385: gap of unknown length
* 36386 38410: contig of 2025 bp in length
* 38411 38510: gap of unknown length
* 38511 39931: contig of 1421 bp in length
* 39932 40031: gap of unknown length
* 40032 42488: contig of 2457 bp in length
* 42489 42588: gap of unknown length
* 42589 44562: contig of 1974 bp in length
* 44563 44662: gap of unknown length
* 44663 47260: contig of 2398 bp in length
* 47261 47360: gap of unknown length
* 47361 49286: contig of 1926 bp in length
* 49287 49386: gap of unknown length
* 49387 51557: contig of 2171 bp in length
* 51558 51657: gap of unknown length
* 51658 54101: contig of 2444 bp in length
* 54102 54201: gap of unknown length
* 54202 56305: contig of 2104 bp in length
* 56306 56405: gap of unknown length
* 56406 59008: contig of 2603 bp in length

59009 59108: gap of unknown length
* 59109 61501: contig of 2393 bp in length
* 61502 61601: gap of unknown length
* 61602 63740: contig of 2139 bp in length
* 63741 63840: gap of unknown length
* 63841 67047: contig of 3207 bp in length
* 67048 67147: gap of unknown length
* 67148 70060: contig of 2913 bp in length
* 70061 70160: gap of unknown length
* 70161 74802: contig of 4642 bp in length
* 74803 74902: gap of unknown length
* 74903 77359: contig of 2457 bp in length
* 77360 77459: gap of unknown length
* 77460 80121: contig of 2662 bp in length
* 80122 80221: gap of unknown length
* 80222 82487: contig of 2266 bp in length
* 82488 82587: gap of unknown length
* 82588 87161: contig of 4574 bp in length
* 87162 87261: gap of unknown length
* 87262 89779: contig of 2518 bp in length
* 89780 89879: gap of unknown length
* 89880 93959: contig of 4080 bp in length
* 93960 94059: gap of unknown length
* 94060 99749: contig of 5690 bp in length
* 99750 99849: gap of unknown length
* 99850 105428: contig of 5579 bp in length
* 105429 105528: gap of unknown length
* 105529 108647: contig of 3119 bp in length
* 108648 108748: gap of unknown length
* 108749 111451: contig of 2704 bp in length
* 111452 111551: gap of unknown length
* 111552 117443: contig of 5892 bp in length
* 117444 117543: gap of unknown length
* 117544 121850: contig of 4307 bp in length
* 121851 121950: gap of unknown length
* 121951 128031: contig of 6081 bp in length
* 128032 128131: gap of unknown length
* 128132 132701: contig of 4570 bp in length
* 132702 132801: gap of unknown length
* 132802 137953: contig of 5152 bp in length
* 137954 138053: gap of unknown length
* 138054 146729: contig of 8676 bp in length
* 146730 146829: gap of unknown length
* 146830 152296: contig of 5467 bp in length
* 152297 152396: gap of unknown length
* 152397 157499: contig of 5103 bp in length
* 157500 157599: gap of unknown length
* 157600 163449: contig of 5850 bp in length
* 163450 163549: gap of unknown length
* 163550 168124: contig of 4575 bp in length
* 168125 168224: gap of unknown length
* 168225 178218: contig of 9894 bp in length
* 178119 178218: gap of unknown length
* 178219 189320: contig of 11102 bp in length
* 189321 189420: gap of unknown length
* 189421 189421: gap of unknown length
* 189422 196694: contig of 7274 bp in length
* 196695 196794: gap of unknown length
* 196795 206464: contig of 9670 bp in length
* 206465 206564: gap of unknown length
* 206565 217803: contig of 11239 bp in length
* 217804 217903: gap of unknown length
* 217904 230534: contig of 12631 bp in length.

FEATURES
source

1. 230534
Location/Qualifiers

/organism="Mus musculus"
/db_xref="taxon:10090"

/clone="RP23-8655"

BASE COUNT 61646 a 50860 c 50276 g 61304 t 6448 others

ORIGIN

Query Match

9.6%; Score 283.6; DB 65; Length 230534;

Best Local Similarity 57.8%; Pred. No. 4.2e-35;
Matches 580; Conservative 0; Mismatches 394; Indels 30; Gaps 3;

Search completed: February 22, 2001, 03:04:53
Job time: 14351 sec

```
OY 821 cccgtgacccgcccccttccagttgtctgttggcaagccttcggcagcaagcccaact 880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221303 CAGGCGAGAGACCGTTTCGTGTGTGGAGTGTGGCCGCGACGTTTCAGCCGACGCTTCACCC 221362

OY 881 tgatcgtccaccgcccgtgtgcaacagggcgagcccccacagtgccccagttgcggga 940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221363 TACTCGGCGACACAGCTACAGCATGAGAGAGAGGCGCTTCGTGTGGCGGAGCTGTGGCC 221422

OY 941 agcgctttaaccaataagccctatctgaacttcgaccgcgcaatccacacggcgagaagc 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221423 AAGGCTTGTTGGCGACGCGCCCGCTGGAGAGACACCGGCGGTGTACACAGCGGCGAGCACG 221482

OY 1001 cctaccggtgcaaaagttgcccgcgcgcttcggcaacaaccacgctgtctcaca 1060
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221483 CGTTCCGCTGCGACAGAGTGTGGAGCTTCGGGAGCGCTTCACACTTACTGCAAGCACCC 221542

OY 1061 gcaagattcacaagcgtatccgaggggttcgcccagggcgcccccgccccggagagcccc 1120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221543 AGCGCATCCACGCGGATCCGCTGTGTCACAGCGGCATTCCTGCGCCCTGCGCGGAGTCC 221602

OY 1121 agctgcagccgcccaggaagttccgcccagcccaaccccgcgtaaccttgaac 1180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221603 CGGACCGCCCGGGGCC-----TTCCATGACAGCGAGTCTGCGAGAG 221645

OY 1181 cggcccaggaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221646 CTTCCCGACGCGGCC-----GTGCTGTAGAGCACCAGCGGCTGCACACCGGAGACA 221698

OY 1241 cccctccctctacagctgagcagactgagcaggaagcttcgcgctgagcgtctcctgc 1300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221699 AGTC-----CTTCGAGATCGGTGAGATGGCGGAGCGCTTCGCGCGCGTTCGGTGTGT 221752

OY 1301 gggcccaccagcgacacgacacgacgagggagccttcaactgcgcgaagtgcgggaaga 1360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221753 TGCACACCGCGCGCGCTGTCACACGCGGCGGAAAGCGCCCTTCGCTGTGCGCGAGTCCGGCGCAA 221812

OY 1361 acttcggcaagaagacgacgacgctggtgagcagctgcgcgtgacatccggcgagagccct 1420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221813 GCTTCGCGGAGCGCTTAATCTACCCAGCACACGCGCATCCACACTGGAGAGCGGCGCT 221872

OY 1421 tgcgctgaggaagtgcgcgcgcgcttcccaaggagccatctgcygcgcacatcgcgc 1480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221873 TCGCCTGGCGCGCAATGTGGCAAGACCTTCGCGAGGCGCCACGGTCAAGGCAACCTGC 221932

OY 1481 cggaccagcccccgatcggccctctgtgtgttcggaactgaggaagccttcgcgcaca 1540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221933 GTGTGCAATACGGGTGAGAAAGCCCTTCGCGTGCCTCCGAGTGCAGCGCCATTACGCCAAGC 221992

OY 1541 aacctactggtggcgacacgcgacacgacacgacacgacacgacacgacacgacacgacac 1600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221993 GACTAAGCTACGCGGCGCAACCAAGACCATATGCGAGAGCCCTTACTGTGCGGTG 222052

OY 1601 actgcgcaaaagccttcagccagaaagttcaactgtgtgcgacggcgacatccacaagc 1660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222053 AGTGGGCGCTGCGCTTCAAGCGAGGTGTGCGGCTCACGAGCACGCGCATCCACACAG 222112

OY 1661 gcgagcgccctagcgctgtccgacgtgcagcgcaagcttcagcagaagttccaactca 1720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222113 GTGACGCGCGCTTCGCTGTCCGATGTGGCAGAGTTTCGCGAGGCGACGCAACCTCA 222172

OY 1721 tcaaccagcgcaagagcaacatccgggacgagccttctgtgtgcatctgtggcaga 1780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222173 CCGACACCGCGCGCATTCATTAAGGCGAGCGACCTTACGCGTGTCCGATGTGGCAAG 222232

OY 1781 ccttcgacgagcagagactcctgcccacacagaagaagcac 1824
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222233 CCTTTCGCGAGGCGCGCAAGCTTCAGGACGACACTGCGACTCAC 222276
```

GenCore Version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2001, 03:04:53 ; Search time 7860.64 Seconds
(without alignments)
246.101 Million cell updates/sec

Title: US-09-477-392-3

Perfect score: 378

Sequence: 1 ggtgagatgcgtcaccg.....tacctgaaccgcccag 378

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 255875100 residues

tal number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_ph:*
- 6: gb_pl1:*
- 7: gb_pl2:*
- 8: gb_pr1:*
- 9: gb_pr2:*
- 10: gb_pr3:*
- 11: gb_ro:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: em_fun:*
- 15: em_hum1:*
- 16: em_hum2:*
- 17: em_in:*
- 18: em_om:*
- 19: em_or:*
- 20: em_ov:*
- 21: em_pat:*
- 22: em_ph:*
- 23: em_pl:*
- 24: em_ro:*
- 25: em_sy:*
- 26: em_sy:*
- 27: em_un:*
- 28: em_vi:*
- 29: gb_htg1:*
- 30: gb_htg2:*
- 31: gb_in1:*
- 32: gb_in2:*
- 33: em_ba1:*
- 34: em_ba2:*
- 35: em_hum3:*
- 36: em_hum4:*
- 37: gb_pr4:*
- 38: gb_htg3:*
- 39: gb_htg4:*
- 40: gb_htg5:*
- 41: gb_htg6:*
- 42: gb_htg7:*
- 43: em_htg1:*

44: em_htg2:*

45: em_htg3:*

46: em_hum5:*

47: gb_pl3:*

48: gb_pr5:*

49: gb_htg8:*

50: gb_htg9:*

51: gb_htg10:*

52: gb_htg11:*

53: gb_htg12:*

54: gb_htg13:*

55: gb_htg14:*

56: gb_in3:*

57: gb_htg15:*

58: gb_htg16:*

59: gb_htg17:*

60: em_htg4:*

61: em_htg5:*

62: em_htg6:*

63: em_htg7:*

64: em_hum6:*

65: gb_htg18:*

66: gb_htg19:*

67: gb_htg20:*

68: gb_htg21:*

69: gb_htg22:*

70: gb_htg23:*

71: gb_v11:*

72: gb_v12:*

73: gb_ba3:*

74: em_htg8:*

75: em_htg9:*

76: em_htg10:*

77: gb_pr6:*

78: gb_pr7:*

79: gb_sts1:*

80: gb_sts2:*

81: gb_pat1:*

82: gb_pat2:*

83: em_htg0:*

84: gb_htg24:*

85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	378	100.0	2878	37 AK025356	AK025356 Homo sapi
2	378	100.0	2878	77 HSA245553	AJ245553 Homo sapi
3	378	100.0	2954	37 AF201303	AF201303 Homo sapi
4	378	100.0	132150	9 AC005586	AC005586 Homo sapi
5	243	64.3	201458	40 AC015887	AC015887 Mus muscu
6	142.6	37.7	759	78 HSM802596	AL162065 Homo sapi
7	126.6	33.5	210923	39 AC004877	AC004877 Homo sapi
8	126.6	33.3	2094	11 MMZRP29	AC013273 Homo sapi
9	126	33.3	2094	11 MMZRP29	X55126 M. musculus
10	125.4	33.2	162570	58 AC073314	AC073314 Homo sapi
11	122.2	32.3	5627	8 AB011129	AB011129 Homo sapi
12	122.2	32.3	154218	10 AC020663	AC020663 Homo sapi
13	120.2	31.8	138278	84 HSI599615	AL109966 Homo sapi
14	118.8	31.4	29802	29 AC007518	AC007518 Homo sapi
15	117	31.0	201458	8 AC015887	AC015887 Mus muscu
16	116.8	30.9	6045	8 AB002324	AB002324 Human mRN
17	116.8	30.9	185306	39 AC013570	AC013570 Homo sapi
18	116.6	30.8	130067	9 AC007228	AC007228 Homo sapi
19	116.6	30.7	1554	37 AF269249	AF269249 Homo sapi
20	115.2	30.5	2330	37 AK023989	AK023989 Homo sapi
21	115.2	30.5	3000	10 AF060503	AF060503 Homo sapi

MEDLINE
 99063792
 2 (bases 1 to 132150)
 Stoneking,T., Ozersky,P., Wohldmann,P. and Le,T.
 TITLE
 The sequence of Homo sapiens PAC clone RP4-584D14
 JOURNAL
 Unpublished
 3 (bases 1 to 132150)
 Waterston,R.H.
 TITLE
 Direct Submission
 Submitted (01-SEP-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 132150)
 Waterston,R.H.
 TITLE
 Direct Submission
 Submitted (12-JUN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 132150)
 Waterston,R.
 TITLE
 Direct Submission
 Submitted (30-SEP-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jun 12, 2000 this sequence version replaced gi:3907529.
 COMMENT

 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@watscn.wustl.edu

 Summary Statistics

 Center project name: H_DJ0584D14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/GRN/CHR7>, send
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP5-820A21, 200 base pair overlap. Actual start of this clone is at base position 1 of RP4-584D14; actual end is at base position 131954 of RP4-584D14.

The sequence RP4-584D14 from base position 21655 to 22775 is GT rich. This region was sized with PCR from genomic DNA and the hindIII digest with band size 3643 real, and 3642 insilico. The

FEATURES	sequence RP4-584D14 from base position 22628 to 22637 is represented by sequence derived by PCR from genomic DNA.
source	Location/Qualifiers 1..132150 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7q31-q35" /clone="RP4-584D14" /clone_lib="RPCI-4" 230..315 /rpt_family="MIR" 960..1288 /rpt_family="MER4-group" 1360..1468 /rpt_family="L2" 1562..1616 /rpt_family="L2" 1733..2080 /note="similar to EST AW510685 (NID:g7148763) hc89e08.x1" 2051..2114 /rpt_family="L2" 2148..3224 /note="CpG island (%GC=70.7, o/e=0.90, #CpGs=107)" 2347..2426 /rpt_family="GC-rich" 3141..3232 /rpt_family="Alu" 3235..3368 /rpt_family="Alu" 3795..3888 /rpt_family="MIR" 4037..4199 /rpt_family="Alu" 4308..4593 /rpt_family="MER4-group" 4872..5309 /note="similar to EST AI827463 (NID:g5448134) wf29f03.x1" 4882..4909 /rpt_family="AT-rich" 5679..5773 /rpt_family="MIR" 5947..5952 /note="match to EST AA482420 (NID:g2210098) zt34f02.r1" 6148..6450 /rpt_family="Alu" 6453..6690 /rpt_family="MER4-group" 6695..7007 /rpt_family="Alu" 7008..7117 /rpt_family="MER4-group" 7180..7355 /rpt_family="MER4-group" 7615 /note="match to EST A1144211 (NID:g3666020) qb93g11.x1" 7883..8397 /note="match to EST W72943 (NID:g1383235) zd54f12.r1" 7886..8357 /note="match to EST A1144211 (NID:g3666020) qb93g11.x1" 7886..8562 /note="similar to EST AW411281 (NID:g6936822) fh11d05.y1" 7886..8403 /note="similar to EST AW246540 (NID:g6589533)" 7886..8451 /note="match to EST AA129215 (NID:g1689084) zn84b02.s1" 7886..8133 /db_xref="gi:1131717" 7886..8146 /note="similar to EST A1866891 (NID:g5542895) wc49f12.x1" 7888..8340 /note="match to EST H37790 (NID:g907289) yp46h05.s1" 7896..8372

misc_feature	/note="assembly_fragment" 7240. .99150
misc_feature	/note="assembly_fragment" 99251. .123995
misc_feature	/note="assembly_fragment" 130096. .163192
misc_feature	/note="assembly_fragment" 163293. .201458
BASE COUNT	5496 a 46124 c 45270 g 5366 t 1902 others
ORIGIN	
Query Match	64.3%; Score 243; DB 40; Length 201458;
Best Local Similarity	83.4%; Pred. No. 1.1e-34;
Matches 276; Conservative	0; Mismatches 55; Indels 0; Gaps 0;
QY	1 ggtgagatgcccgtgacccgccccttcagctgtgtcgtgtgtgtgagcaagcgttcggac 60
Db 132529	GGTGGAGAGCGCTGTGGACGCCCTTCCAGTGTGCTGTGGCGCAAGCGTTCCGCC 132588
QY	61 aagcccaactatcgctacccgcgcgtgcacacggagcgagcgagccaccagtacccc 120
Db 132589	AAGCCCAACTATCGCTACCCGCCCGCGTGCACACTGGTGAGCGACACACCAAGTCCCA 132648
QY	121 gsgtgcgggaagcgcttccacaataagccctatctgacttgcacgcgcgcacacc 180
Db 132649	GAGTGGGGAACGCTTACCAACAAGCCCTACTGTGACTCGACCGCGCATATAC 132708
QY	181 ggcgggaagccttaccctgctgcaaaagtgcgcgccgccttcgcgcaacaaccacgtg 240
Db 132709	GGCGGAACCCCTACCCACGACAGTGTGGCCGCCGCTTCCGCCAACAACCACTG 132768
QY	241 cgtctcacagaagaattcacagaagatccggagggtgcggccagcgcccccggcccg 300
Db 132769	TGTGTGCACAGCAAAATCCACAAGCGCTTGGAGGTCTCAGCCGAGGCTGTCCACACCCC 132828
QY	301 gggagcccccagctgcacgcgcgcgcgcag 331
Db 132829	GAGAGTCACACGATTCGACGACGAGCGCTTAGG 132859
RESULT 6	
LOCUS	HSMB02596/1 759 bp mRNA PRI 23-MAR-2000
DEFINITION	Homo sapiens mRNA; cDNA DKFZp762K135 (from clone DKFZp762K135).
ACCESSION	AL162065
VERSION	AL162065.1 GI:7328138
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 759)
JOURNAL	Bloecher,H., Bloecher,M., Brandt,P., Mewes,H.W., Weil,B. and Wiemann,S.
COMMENT	Direct Submission Submitted (15-MAR-2000) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp762K135) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES	Location/Qualifiers
SOURCE	1. .759
	/organism="Homo sapiens"
	/db_xref="taxon:9606"

repeat_region /rpt_family="Retroviral"
12617. .12755
/rpt_family="L1"
12768. .12890
/rpt_family="MIR"
12891. .13196
/rpt_family="Alu"
13197. .13209
/rpt_family="MIR"
13314. .13618
/rpt_family="Alu"
13693. .14004
/rpt_family="Alu"
14040. .14162
/rpt_family="MIR"
15625. .15931
/rpt_family="Alu"
16550. .16702
/rpt_family="L1"
18090. .19036
/rpt_family="L1"
19479. .19798
/note="match to EST AA939299 (NID:g3099212) ol78g08.s1"
misc_feature 19479. .19656
/note="similar to EST H89462 (NID:g1079892)"
19817. .20311
/note="match to EST AA625192 (NID:g2537577) af67h06.r1"
19860. .21428
/note="CPG island (%C=70.4, o/e=0.70, #CGs=171)"
misc_feature complement(19861. .20173)
/note="match to EST AA923119 (NID:g3070428) ok91e12.s1"
19861. .20239
/note="similar to EST AA923119 (NID:g3070428) ok91e12.s1"
misc_feature complement(19898. .20090)
/note="match to EST AA771702 (NID:g2823485) ai32a10.s1"
misc_feature complement(19937. .20235)
/note="match to EST AA885789 (NID:g3000897) oj35a06.s1"
gene complement(20009. .21523)
/gene="W0GSC:H.DJ0751H13.3"
complement(20009. .21523)
/gene="W0GSC:H.DJ0751H13.3"
/note="zinc finger-like; similar to P52742 (PID:g1731411):"
H.DJ0751H13.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC36300.1"
/db_xref="GI:3638956"
/translation="MIRKRVDEDEQEAEEVEWPHQLSLPSPAPADLGHLLAAAY
LPPAGALSGALSGWGPMPKPYGCGCEERFPDOLLRLHNRHLEBGPACPD
GSGFROAMLHSHRSHGEPFPCSEDDRKSKAHLTRLRHTEGRRYQACGARS
RSQKILHSHQKTHGRRPPTCEKRRFKTKLLTHQRHHTGERYQACGARS
TIKQHLVRRQRVHTAGRRRSPSSASHSTABSP17SFEPKPFACSDGLSGW
KNLATHQICDCRGRPFQDECALGATVDAPAKLASAFPGGPGSDGVQORAR
SESRFECCHDCEGSHQOHLARPRVHTGERPFACQCDRRFGSPMLVAHSHASR
AAPFACQCGSRFSRSHLGRHOAVHTGSRPHACVCAFRSSKTNLVROAHTGSR
PSPCCGCGSFRKTHLVROHIGEEAHAAPADALAPAMSAPEVAPPLPF"
misc_feature complement(20014. .20772)
/gene="W0GSC:H.DJ0751H13.3"
/note="match to EST AA771702 (NID:g2823485) ai32a10.s1"
misc_feature complement(20059. .20239)
/gene="W0GSC:H.DJ0751H13.3"
/note="match to EST AA923119 (NID:g3070428) ok91e12.s1"
misc_feature complement(20097. .20319)
/gene="W0GSC:H.DJ0751H13.3"
/note="match to EST AA885789 (NID:g3000897) oj35a06.s1"
20176. .20322
/note="match to EST AA894411 (NID:g3030812) oe79b04.s1"
20298. .20740
/note="similar to Mus musculus EST AA021873 (NID:g1485629)
mb85f12.r1"
20534. .20724
/note="match to EST AA044857 (NID:g1523196) zf54d12.r1"
20819. .21219

```

repeat_region /note="match to EST AA757828 (NID:g2805651) zg44d02.s1
22351. .22486 /rpt_family="MIR"
repeat_region 22848. .23146 /rpt_family="Alu"
repeat_region 23679. .23794 /rpt_family="MIR"
repeat_region 23773. .23846 /rpt_family="L2"
repeat_region 24894. .24982 /rpt_family="L2"
repeat_region 24983. .25271 /rpt_family="Alu"
repeat_region 25272. .25616 /rpt_family="L2"
misc_feature 26993. .29260 /note="CpC_island (MGC-73.8, o/e=0.90, #CpGs=195)"

```

Query Match	33.5%;	Score 126.6;	DB 8;	Length 128361;
Best Local Similarity	67.8%;	Pred. No. 3.7e-14;		
Matches 177; Conservative	0;	Mismatches 84;	Indels 0;	Gaps 0;

QY	11	ccctgagacgcgccttcacagtgctcgtttgtgaaagcgttcctcgagacaagcccaact	70
Db	21081	CGGGCGACGGGCGCTACCCCGTGCGGAGATGCGGACAGCGCTTACGCCAGAAATCCACC	21022
QY	71	tgatgcgtcacccgcgcgtgtgacacacgagcgagcgagcccccacagtgcccgcagtgcygga	130
Db	21021	TGGGCTCGCACCAAAAGACCCACACCGCGCGAGCGCCCTTCCCTCGACAGGAAATCGAGA	20962
QY	131	agcgcttaccgaataagaacctatcgtactgcagccggcgatcccaaccgcggcgagaagc	190
Db	20961	AGCGCTTTCCGACAGACAGCGACATTGATTGGCGACCAAGCGCATTCATACAGGGCGAGAAGC	20902
QY	191	ccctaccggtlgcaaaagatgycgcgcgcgtcttcgycgcaaaacccaacctgtcgttcaca	250
Db	20901	CCCTACCAATGGCGACAGTGGCGACGACACTTCACACACAAAGAGCATTTGGTGGCGCAC	20842
QY	251	gcaagattcacagaagcatcgc	271
Db	20841	AAAGGTGCACCAAGCGCGC	20821

RESULT..	8
AC013273/c	
LOCUS	AC013273 210923 bp DNA HTG 15-JUL-2000
DEFINITION	Homo sapiens chromosome 7 clone RP11-428D5, WORKING DRAFT SEQUENCE,
ACCESSION	AC013273
VERSION	10 unordered pieces.
KEYWORDS	AC013273.5 GI:2111421
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT. human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 210923) Waterston,R.H.
TITLE	The sequence of Homo sapiens clone unpublished
JOURNAL	2 (bases 1 to 210923)
REFERENCE	Waterston,R.H.
AUTHORS	Direct Submission
TITLE	Submitted (05-NOV-1999) Genome Sequencing Center, Washington
JOURNAL	MO 63108 USA University School of Medicine, 4444 Forest Park Parkway, St. Louis, On Jul15, 2000 this sequence version replaced gi:8990985.
COMMENT	

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 ----- Project Information -----
 Center project name: H_NH0428D05

----- Summary Statistics -----

Sequencing vector: M13: 74%
 Chemistry: Dye-primer ET: 74% of reads
 Chemistry: Dye-terminator Big Dye: 26% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 200959 bases at least Q40
 Consensus quality: 204292 bases at least Q30
 Consensus quality: 206312 bases at least Q20
 Insert size: 203000; agarose-IP
 Insert size: 210023; sum-of-ctrls
 Quality coverage: 4.85 in Q20 bases; agarose-IP
 Quality coverage: 4.93 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2429: contig of 2429 bp in length
 * 2430 2529: gap of unknown length
 * 2530 4617: contig of 2088 bp in length
 * 4618 4717: gap of unknown length
 * 4718 11807: contig of 7090 bp in length
 * 11808 11907: gap of unknown length
 * 11908 20564: contig of 8557 bp in length
 * 20565 29298: gap of unknown length
 * 29299 29398: contig of 8734 bp in length
 * 29399 45700: gap of unknown length
 * 45701 45800: contig of 16302 bp in length
 * 45801 63662: gap of unknown length
 * 63663 63762: gap of unknown length
 * 63763 104507: contig of 40745 bp in length
 * 104508 104607: gap of unknown length
 * 104608 150643: contig of 46036 bp in length
 * 150644 210923: gap of unknown length
 * 210923 210923: contig of 60180 bp in length.

FEATURES

source

1. 210923
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /clone="RP11-428D5"
 1. 2429
 /note="assembly_name:Contig8"
 misc_feature 2530..4617
 /note="assembly_name:Contig9"
 misc_feature 4718..11807
 /note="assembly_name:Contig10"
 misc_feature 11908..20464
 /note="assembly_name:Contig11"
 misc_feature 20565..29298
 /note="assembly_name:Contig12"
 /clone_end="7"
 vector_side:right"
 /note="assembly_name:Contig13"
 misc_feature 29399..45700
 /note="assembly_name:Contig14"
 misc_feature 45801..63662
 /note="assembly_name:Contig15"
 vector_side:left"
 /clone_end="SP6"
 63763..104507
 /note="assembly_name:Contig16"
 misc_feature 104608..150643
 /note="assembly_name:Contig17"
 misc_feature 150744..210923
 /note="assembly_name:Contig17"
 BASE COUNT 54201 a 51422 c 51731 g 52663 t 906 others
 ORIGIN

Query Match 33.5%; Score 126.6; DB 39; Length 210923;
 Best Local Similarity 67.8%; Pred. No. 3.1e-14;
 Matches 177; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 11 ccgcgcacccccccttcagtgctgctgtgtgcaagcgttcgcgcacacccacact 70
 Db 98278 CGGCGAGCGGCGCTTACCGCGCGGAGTGCGGCAAGCGCTTACGACAGATCCACC 98219
 QY 71 tgatcgtcacccgcgcgtgtgacacagcgagcgcccccacagtcgagtgcgga 130
 Db 98218 TGGGCTCGCACCAAAACACCCACGCGCGCGCGCTTCCCTTCGACGGAATGCCAGA 98159
 QY 131 agcgccttaccataagccctatctgacttcgcgcgcgcgcacacacccgcgcagaagc 190
 Db 98158 AGCGCTTCCGCAAGAAAGACACACTTGTGGGCACGACGCATCATATGCGGAGAGGC 98099
 QY 191 cctaccgttcaagaagtggcgccgcttcgcgcgcgcgcacacacccacactgtctaca 250
 Db 98098 CCTRACAGTCCGCACTGCGCACGCTTCACGACAAAGCAGCACTGTGCGGACC 98039
 QY 251 gcaagattcacacgcgattccg 271
 Db 98038 AAAGGTCACACGACGCGCG 98018

RESULT 9

LOCUS MZFP29 2094 bp mRNA ROD 13-APR-1992
 DEFINITION M.musculus zfp-29 gene for zinc finger protein.
 ACCESSION X55126.1 GI:55470
 VERSION X55126.1 GI:55470
 KEYWORDS spermatogenesis; zfp-29 gene; zinc-finger.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2094)
 AUTHORS Rodentia; Sciuognath; Muridae; Murinae; Mus.
 TITLE Direct Submission

JOURNAL Submitted (03-NOV-1990) Denny P., Institute of Cancer Research,
 Denny P.
 REFERENCE 2 (bases 1 to 2094)
 AUTHORS Chester Beatty Laboratories, 237 Fulham Road, LONDON, SW3 6JB, UK
 TITLE A zinc finger protein-encoding gene expressed in the post-meiotic
 phase of spermatogenesis

JOURNAL Gene 106 (2), 221-227 (1991)
 MEDLINE 92039080
 FEATURES

source

Location/Qualifiers

1. 2094
 /organism="Mus musculus"
 /strain="ssp. domesticus"
 /db_xref="taxon:10090"
 /dev_stage="adult male"
 /tissue_type="testis"
 /clone_lib="Lambda ZAP cDNA"
 104..1948
 /gene="zfp-29"
 104..1948
 /gene="zfp-29"
 /codon_start=1
 /protein_id="CAA38920.1"
 /db_xref="GI:55471"
 /translation="MAAYPAVSTPSPIVOYQPEBDAQAEVYTMILLEDANVQAVL
 QEDGPSEFPPOSAGSGSPQEDBAAGCPGALVRFELCRMLRPVHTKQMTVLP
 KETIQALQHPREPSSEAAVALVDLTQTRHSDFEIQSENGENSNEDMEGVESHMF
 LNTSGEGGQSDPDSDFERDVGSGGAQGHAGDEPRVYPSREVGQDILGQGYLG
 EKPYECPOGKTFSRKSLITHRTHTGKYYKDCDCKSGFSQSNFSRHTHTGK
 PKYCRDCKSFSRSANLITHRTHTGKFPQCAEKGSRSPSNLIARHTHTGKPY
 SCPECKSFRGNSLNTHTGKTHGKPYACKCKGSEFSYNSNLIARHTHTGKPKY
 TECGKFSQSSALITHRTHTGKPYQCEGCKNFSRSSNLATHRTHTLVERPKGL


```

RESULT 11
AB011129
LOCUS      Homo sapiens mRNA          PRI          10-Apr-1998
DEFINITION Homo sapiens mRNA for KIAA0557 protein, partial cds.
ACCESSION AB011129
VERSION    AB011129.1 GI:3043637
KEYWORDS   KIAA0557 protein.
SOURCE     Homo sapiens male brain cDNA to mRNA, clone_lib:pbuescriptlii SK
           plus clone:HH1334.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 5627)
AUTHORS   Ohara,O., Nagase,T. and Ishikawa,K.
JOURNAL   Direct Submission
TITLE     Submitted (13-FEB-1998) to the DDBJ/EMBL/GenBank databases. Osamu
           Ohara, Kazusa DNA Research Institute, DNA Technology; 1532-3 Yana,
           Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@foe.kazusa.or.jp,
           Tel:+81-438-52-3913, Fax:+81-438-52-3914)
REFERENCE  2 (stiles)
AUTHORS   Nagase,T., Ishikawa,K., Miyajima,N., Tanaka,A., Kotani,H.,
           Nomura,N. and Ohara,O.
JOURNAL   Prediction of the coding sequences of unidentified human genes. IX.
           The complete sequences of 100 new cDNA clones from brain which can
           code for large proteins in vitro
TITLE     DNA Res. 5 (1), 31-39 (1998)
FEATURES   98290545
            location/Qualifiers
            1..5627
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="HH1334"
               /clone_lib="pbuescriptlii SK plus"
               /sex="male"
               /tissue_type="brain"
               1..1483
                  /gene="KIAA0557"
                  <1..1483
                     /gene="KIAA0557"
                     /codon_start=2
                     /product="KIAA0557 protein"
                     /protein_id="BAA25483.1"
                     /db_xref="GI:3043638"
                     /translation="TTRPEKGGPPEAMATVYGLQPTLEQDLEQELLIVKPEEP
                     CLEEPSVETEDBSPETFRQLFRLCTQVAGPREALSRMLELCCNWRPELRKEQD
                     LELLVGLQFLIVLPEGLQIARVREQQPSGEAVLVGLQKPKRHQROSELLSDD
                     VPLIDGQFLFKHQEAQEDQSLDEEERAFSSQDPPADLSHPQRPILWREGPPAPR
                     HOEWAASPELSANQVAVNLEDAVAVLLEGGPECMPPADADALENGPQIQLEGGPQ
                     DGRDADALRMENYRVLSARCOGPGHPPLPGQAPAVRGLIVRPDQRPQPPGRARRSHQAN
                     DKPTCPDEGKGSGKTSHLTRKHQTHNGHGERYKLVGKSGKSDSNSTHQRVITGKE
                     RPYCPGCGKRFSOSSSLVILKTHRSGERPACTCGKRFNNSHFSARRKTHTEKPT
                     TPCPGGGRFGRGTDLNRHQRTHMAGSLPLQPVAPGPGAKA"
BASE COUNT      1173 a      1606 c      1690 g      1158 t
ORIGIN
Query Match      32.3%; Score 122.2; DB 8; Length 5627;
Best Local Similarity 60.1%; Pred. No. 7.6e-13;
Matches 223; Conservative 0; Mismatches 143; Indels 5; Gaps 1;

0y      8 atgccgtgacgcgcctctcagtgctgctgtgtggcaagagcttcgscgcaagccca 67
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      996 ATGGGCTGACAAAGCCGTACACCTGCCCCGAATGTGGCAAAAGCTTACAGACAGCTCC 1055

0y      68 acttagctcactacgcgcgctgcacacagcgcgcaagcgccacagctgcccagatgcg 127
        ||||| ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| |||
Db      1056 ACTTGACCAAGACACAGCGCACACACAGAGGGGAGCGGCTTACAAAGTGCCATGCTGTG 1115

0y      128 ggaagcgcttaccacaaataagccctatctgtacttcgcacacgcgcacatcacacgcggsaga 187

```

[illegible]

```

repeat_region /rpt_family="Alu" 3187. .3504
repeat_region /rpt_family="Alu" complement(3571. .3890)
repeat_region /rpt_family="Alu" complement(3859. .4163)
repeat_region /rpt_family="Alu" 4198. .4503
repeat_region /rpt_family="Alu" 4826. .5155
repeat_region /rpt_family="Alu" 5749. .6045
misc_feature /rpt_family="Alu" complement(6848. .6983)
repeat_region /note="GRAIL 2 excellent exon, frame 1" 7473. .7776
repeat_region /rpt_family="Alu" 7931. .8385
repeat_region /rpt_family="Alu" 8808. .9139
repeat_region /rpt_family="Alu" 9394. .9682
repeat_region /rpt_family="Alu" 9515. .9812
repeat_region /rpt_family="Alu" 9702. .9987
repeat_region /rpt_family="Alu" 9965. .10332
repeat_region /rpt_family="Alu" complement(10403. .10835)
repeat_region /rpt_family="Alu" complement(10801. .10977)
repeat_region /rpt_family="Alu" 11580. .11780
repeat_region /rpt_family="MER42" 11803. .12101
repeat_region /rpt_family="Alu" 12282. .12576
repeat_region /rpt_family="Alu" 12643. .13073
repeat_region /rpt_family="Alu" 12923. .13237
repeat_region /rpt_family="Alu" 13420. .13725
repeat_region /rpt_family="Alu" 13761. .14148
repeat_region /rpt_family="Alu" 14072. .14371
repeat_region /rpt_family="Alu" complement(14401. .14696)
repeat_region /rpt_family="Alu" 15061. .15366
repeat_region /rpt_family="Alu" 15383. .15706
repeat_region /rpt_family="Alu" complement(16083. .16185)
repeat_region /rpt_family="L1" complement(16192. .16347)
repeat_region /rpt_family="Alu" 16559. .16831
repeat_region /rpt_family="Alu" complement(18420. .18718)
repeat_region /rpt_family="Alu" complement(18518. .18917)
repeat_region /rpt_family="Alu" complement(18754. .19036)
repeat_region /rpt_family="Alu" complement(19456. .19559)
misc_feature /note="GRAIL 2 excellent exon, frame 1" 20239. .20397
repeat_region /rpt_family="MER20" complement(20808. .21104)
repeat_region /rpt_family="Alu"

```

```

repeat_region complement(22116. .22443)
misc_feature /rpt_family="Alu" complement(22789. .22948)
repeat_region /note="GRAIL 2 excellent exon, frame 1" 23346. .23646
repeat_region /rpt_family="Alu" complement(24046. .24341)
repeat_region /rpt_family="Alu" 25205. .25461
repeat_region /rpt_family="Alu" 25605. .25862
repeat_region /rpt_family="L1" 25914. .25981
repeat_region /rpt_family="Alu" complement(26044. .26355)
repeat_region /rpt_family="Alu" 28437. .28732
repeat_region /rpt_family="Alu" 30597. .30772
misc_feature /note="GRAIL 2 excellent exon, frame 2" 31300. .31655
repeat_region /rpt_family="Alu" 32059. .32124
repeat_region /rpt_family="Alu" complement(32302. .32492)
repeat_region /rpt_family="Alu" 32507. .32839
repeat_region /rpt_family="Alu" 33649. .34040
repeat_region /rpt_family="Alu" complement(34202. .34278)
repeat_region /rpt_family="LTR37B" 34279. .34556
repeat_region /rpt_family="Alu" 35017. .35495
repeat_region /rpt_family="Alu" complement(36639. .36991)
repeat_region /rpt_family="Alu" 37667. .37808
misc_feature /note="GRAIL 2 excellent exon, frame 0" complement(37920. .37961)
repeat_region /rpt_family="MIR" 38430. .38745
repeat_region /rpt_family="Alu" 38789. .38899
repeat_region /rpt_family="MER3" complement(40939. .41089)
repeat_region /rpt_family="Alu" 42294. .42589
repeat_region /rpt_family="Alu" complement(43093. .43396)
repeat_region /rpt_family="Alu" complement(43274. .43695)
repeat_region /rpt_family="Alu" 44405. .44482
repeat_region /rpt_family="MLTIC" 44515. .44693
repeat_region /rpt_family="Alu" 45916. .46254
misc_feature /rpt_family="Alu" complement(46300. .46347)
repeat_region /note="GRAIL 2 excellent exon, frame 1" 46426. .46596
repeat_region /rpt_family="MER20" 47139. .47448
repeat_region /rpt_family="Alu" 47535. .47943
repeat_region /rpt_family="HE1" complement(47939. .48377)
repeat_region /rpt_family="Alu" 48431. .48778
repeat_region /rpt_family="L1" 48877. .49260

```



```

misc_feature      /note="assembly_fragment"
4159. 5831
/note="assembly_fragment"
5932. 7380
misc_feature      /note="assembly_fragment"
7481. 8820
/note="assembly_fragment"
8921. 10519
misc_feature      /note="assembly_fragment"
10620. 13322
/note="assembly_fragment"
13423. 15194
misc_feature      /note="assembly_fragment"
15295. 18405
/note="assembly_fragment"
18506. 20293
misc_feature      /note="assembly_fragment
clone_end:77
vector_side:left"
20394. 24037
misc_feature      /note="assembly_fragment"
24138. 30849
/note="assembly_fragment"
30950. 44871
misc_feature      /note="assembly_fragment"
44972. 61254
/note="assembly_fragment"
61355. 74139
misc_feature      /note="assembly_fragment"
74240. 99150
/note="assembly_fragment"
99251. 129995
misc_feature      /note="assembly_fragment"
130096. 163192
/note="assembly_fragment"
163293. 201458
misc_feature      /note="assembly_fragment"
201458. 24037
BASE COUNT      54496 a 46124 c 45270 g 53666 t 1902 others
ORIGIN
Query Match      31.0%: Score 117; DB 40; Length 201458;
Best Local Similarity 62.5%: Pred. NO. 1.5e-12;
Matches 183; Conservative 0; Mismatches 110; Indels 0; Gaps 0.
QY      4      ggagatgcgcgcgacccgcccctccagatgtgctgtgttggaagcgcttcgcgcaag 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71332 GGGGCTGGCCCGCGCAGCGCCCTTCTGTGTGTCTCCGAGTGGCGCAAGGCTTTCAGCGTCANA 71273
QY      64      cccaatgatcgtcaccgcgcgcgtgtgacacgagcgagcgagccccaacagtgcccgag 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71272 CACAACTGTGAGGTGCACACGACGACACCGGCGAGCGGCCCTTGTGCGCTTCCCGCGAG 71213
QY      124     tgcgggaagcgcttaacacaataagccctatactgacttcgacgcgcgcgcatccacaacgcgc 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71212 TGCGGCCGCTCTTACGCTTCAGGCTCAAGCAGATGTGCTACGCAACACGCGCATCCACAGCGGC 71153
QY      184     gagagagccctaccgctgtgcaagaagatgctggcgccgcgcgtccgcgcacaaacccaactgctg 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71152 GAGAGAGCGCGCACCACTGGCGCAGTGGCGTCTCTTCCCGGACCGCGGCTTCCCTGCTC 71093
QY      244     tctacagacaagatcacaagcgtatccgaggggtgtgcgccacgagcgcccccg 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71092 AACCAACGACGCGCACCAACGCGGCGCATGGCCACGCGCACCGCGCGCTTCCCGG 71040

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2001, 03:09:52 ; Search time 7860.64 Seconds
(without alignments)
115.238 Million cell updates/sec

Title: US-09-477-392-5
Perfect score: 177
Sequence: 1 cgatccggagggtcggccca.....tcgaagccccccctcctc 177

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 111813 seqs, 255875100 residues
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vl:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: em_ba1:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_vil:*
72: gb_vl2:*
73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query length	DB ID	Description
1	177	100.0	2878	37 AK025356	AK025356 Homo sapi
2	177	100.0	2878	77 HS245553	AJ245553 Homo sapi
3	177	100.0	2954	37 AF201303	AF201303 Homo sapi
4	177	100.0	132150	9 AC005586	AC005586 Homo sapi
5	52.4	29.6	166341	39 AC013381	AC013381 Homo sapi
6	52.4	29.6	244410	58 AC069214	AC069214 Homo sapi
7	52.2	29.5	63121	54 AC034263	AC034263 Homo sapi
8	50.2	28.1	78690	54 AC016541	AC016541 Homo sapi
9	49.8	28.1	8058	73 SCE56	AC016541 Homo sapi
10	49.2	27.8	145540	48 AP001052	AL138852 Streptomy
11	49.2	27.8	340000	48 AP001052	AP001052 Homo sapi
12	49.2	27.7	65523	51 AC024266	AC024266 Homo sapi
13	49	27.6	72360	50 AC022984	AC022984 Homo sapi
14	48.8	27.5	68581	53 AC027375	AC027375 Homo sapi
15	48.6	27.3	28826	2 MTY041	AL021958 Mycobacte
16	48.4	27.2	135119	38 AC011578	AC011578 Homo sapi
17	47.8	27.0	63082	50 AC022663	AC022663 Homo sapi
18	47.6	26.9	93821	49 AC021596	AC021596 Homo sapi
19	47.6	26.9	151340	58 AC069132	AC069132 Homo sapi
20	47.2	26.7	340000	77 HS21C102	AL163302 Homo sapi
21					

C	22	47	26.6	2028	47	SS1382828	ALI32828 Spermatzo
C	23	47	26.6	61676	55	AC060800	AC060800 Homo sapi
C	24	47	26.6	208936	38	AC010821	AC010821 Homo sapi
C	25	46.8	26.4	5186	73	SADNAIT2	X87940 S. aurantiac
C	26	46.8	26.4	40901	73	SCC24	ALI33003 Streptomy
C	27	46.8	26.4	159715	65	AC079369	AC079369 Mus muscu
C	28	46.4	26.2	371123	73	SC5F7	AL096872 Streptomy
C	29	46.4	26.2	59978	50	AC022815	AC022815 Homo sapi
C	30	46.4	26.2	109488	77	HS289J15	AL031774 Human DN
C	31	46.4	26.2	193782	65	AC079276	AC079276 Mus muscu
C	32	46.4	26.2	215218	59	AC076968	AC076968 Homo sapi
C	33	46.4	26.2	213661	66	ALI13825	ALI13825 Homo sapi
C	34	46.2	26.1	61676	55	AC060800	AC060800 Homo sapi
C	35	46.2	26.1	136551	55	AC048354	AC048354 Homo sapi
C	36	46.2	26.1	141952	67	ALI161774	ALI161774 Homo sapi
C	37	46.2	26.1	208279	39	AC013564	AC013564 Homo sapi
C	38	46	26.0	54298	65	AC080041	AC080041 Homo sapi
C	39	45.8	25.9	2910	78	HSB63	X69910 H.sapiens p
C	40	45.8	25.9	14844	7	MBU34849	U34849 Mycobacteri
C	41	45.8	25.9	38500	2	MTCY39	Z74025 Mycobacteri
C	42	45.8	25.9	70408	66	AC083993	AC083993 Homo sapi
C	43	45.8	25.9	153559	52	AC025000	AC025000 Homo sapi
C	44	45.8	25.9	160326	59	AC073530	AC073530 Homo sapi
C	45	45.8	25.9	162976	53	AC026302	AC026302 Homo sapi

ALIGNMENTS

		/note="Highly similar to Hs245553 Homo sapiens mRNA for AP4 zinc finger protein"	
BASE COUNT	560 a	974 c	826 g
ORIGIN			518 t
100.0%; Score 177; DB 37; Length 2878;			
Query Match	Best Local Similarity	100.0%;	Pred. No. 1.4e-18;
Matches 177;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0
QY 1	cgatccgagggatcgccgccaagcgcgcgcccccgccgagggagagcccccagctgcacagccggc	60	
Db 970	CGATCCGAGGGGATCGGCCACAGGCCGCCGCCCGCGGAGAGCCGCCACGCTGCACAGCCGGC	1029	
QY 61	ccccgagggatcgcgcgcgcgcgaagccaccccccgcgcggttaaccttgaaacccgcccagaagccg	120	
Db 1030	CCCCGAGGATCGCGCGCGCCGAGGCCACCCCGGGGGTACCTCTTAATAACGGGCCACGAGAGCCG	1089	
QY 121	ccgcacgagggcccgccgacagagcacccgcagagaccgagatgtaagcccccccttccttc	177	
Db 1090	CCGCCAGGGGCCCGCCACAGAGCACCCCGCAGACCCGATGGAAGCCCGCCCTCTCCTC	1146	

```

RESULT      1
AK025356
LOCUS
DEFINITION
  Homo sapiens CDNA: FLJ21703 f1s, clone COL09895, highly similar to
  HS245533 Homo sapiens mRNA for Ap4 zinc finger protein.
ACCESSION
  AK025356
VERSION
  AK025356.1 GI:10437856
KEYWORDS
  oligo capping; f1s (full insert sequence).
SOURCE
  Homo sapiens
  Homo sapiens colon CDNA to mRNA, clone_lib=COL clone:COL09895.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
  1 (stes)
  Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
  Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
  Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  NEDO human cDNA sequencing project
  Unpublished (2000)
  2 (bases 1 to 2878)
  Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
  Shibahara,T., Tanaka,T. and Nakamura,Y.
  Direct Submission
  Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
  Sugano, Institute of Medical Science, University of Tokyo,
  Laboratory of Genome Structure Analysis, Human Genome Center;
  Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
  (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
  Fax:81-3-5445-5416)
COMMENT
  NEDO human cDNA sequencing project supported by Ministry of
  International Trade and Industry of Japan; CDNA full insert
  sequencing: Research Association for Biotechnology; cDNA library
  construction, 5'- 6' 3'-end one pass sequencing: Department of
  Virology and Human genome Center, Institute of Medical Science,
  University of Tokyo (partly supported by Science and Technology
  Agency).
FEATURES
  source
    1..2878
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="COL09895"
    /clone_lib="COL"
    /tissue_type="colon"
    /note="Cloning vector pME18SFL3"
  misc_feature
    1..2878

```

LOCUS	2878	bp	mRNA	PRI	18-AUG-1999
DEFINITION	Homo sapiens mRNA for AP4 zinc finger protein.				
ACCESSION	U224553				
VERSION	U224553.1				
KEYWORDS	AP4 gene; AP4 protein; zinc finger protein.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Dobner,T.G., Fischer,M. and Groitl,P.				
JOURNAL	Cloning of a novel zinc finger protein				
REFERENCE	Unpublished				
TITLE	2 (bases 1 to 2878)				
JOURNAL	Dobner,T.G.				
FEATURES	Direct Submission				
SOURCE	Submitted (17-AUG-1999) Dobner T.G., Institut fuer Medizinische Mikrobiologie und Hygiene, Universitaet Regensburg, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, GERMANY				
gene	1..2878				
CDS	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/cell_line="HeLa"				
	106..1776				
	/gene="AP4"				
	106..1776				
	/gene="AP4"				
	/note="zinc finger protein"				
	/codon_start=1				
	/product="AP4 protein"				
	/protein_id="CAB53100.1"				
	/db_xref="GI:5748565"				
	/translation="MCPAQPRLLSGPQSEPTQLKESRGLRQOGTSVAQSGAQAQPGI				
	AHRCAHCRHHPGFWALVLTIRRCOARLPDCEPCGRGFRAPFLAIRVYHAATPPL				
	LGACHLGQGSFRGVALVCLHRAHSAKAPRIAPCKERRRWRKQLRAHLQROCHPPR				
	PAAPRFICNGCGRGFAQMDLVAAKRVVALEAAKAVLGPRSRGPAVTARPGGQVLA				
	DAADRPQACACCGKFRFRHKPMLIAHRRVHTGERPHOCDEGKRTNKPYLSHRHITH				
	GMDPYPCGEGRRRHKPNLISHKTIKRSRGSQAAPGSGPOLAGPOBSAEPPT				
	AVPLKPKQEPPEPPAPRPHQDPTIAPPPLSLSCDCCGSFLERFLRAHQROHTGPRPP				
	TCACGKCNFSGKTHLVASHSVHSGERFPACGEGRSQSGHLAHRDAPDPPFVOC				
	PDGKAAPRAKPYLAHRRIHITGKDPVPCDGAAPRQKSNLVSRRKIHITGERSYACPP				
	CDRSFSQKSNLTIHRKSHIRDAFCACICGOTFDEERFLAHQKKHDV"				
BASE COUNT	528	a	973	c	839 g 538 t
ORIGIN	100.0%; Score 177; DB 77; Length 2878;				

*	1900	2775:	contlg of 876 bp	in length
*	2756	2875:	gap of 100 bp	in length
*	2876	3771:	contlg of 896 bp	in length
*	3772	3871:	gap of 100 bp	in length
*	3872	4761:	contlg of 890 bp	in length
*	4762	4861:	gap of 100 bp	in length
*	4862	5736:	contlg of 875 bp	in length
*	5737	5836:	gap of 100 bp	in length
*	5837	6640:	contlg of 804 bp	in length
*	6641	6740:	gap of 100 bp	in length
*	6741	7617:	contlg of 877 bp	in length
*	7618	7717:	gap of 100 bp	in length
*	7718	8586:	contlg of 879 bp	in length
*	8597	8696:	gap of 100 bp	in length
*	8697	9555:	contlg of 859 bp	in length
*	9556	9655:	gap of 100 bp	in length
*	9656	10517:	contlg of 862 bp	in length
*	10518	10617:	gap of 100 bp	in length
*	10618	11478:	contlg of 861 bp	in length
*	11479	11578:	gap of 100 bp	in length
*	11579	12420:	contlg of 842 bp	in length
*	12421	12530:	gap of 100 bp	in length
*	12521	13422:	contlg of 902 bp	in length
*	13423	13522:	gap of 100 bp	in length
*	13523	14362:	contlg of 860 bp	in length
*	14383	14482:	gap of 100 bp	in length
*	14483	15347:	contlg of 865 bp	in length
*	15348	15447:	gap of 100 bp	in length
*	15448	16303:	contlg of 856 bp	in length
*	16304	16403:	gap of 100 bp	in length
*	16404	17266:	contlg of 866 bp	in length
*	17270	17369:	gap of 100 bp	in length
*	17370	18239:	contlg of 870 bp	in length
*	18240	18339:	gap of 100 bp	in length
*	18340	19122:	contlg of 853 bp	in length
*	19123	19222:	gap of 100 bp	in length
*	19293	20160:	contlg of 868 bp	in length
*	20161	20260:	gap of 100 bp	in length
*	20261	21120:	contlg of 860 bp	in length
*	21121	21220:	gap of 100 bp	in length
*	21221	22076:	contlg of 856 bp	in length
*	22077	22176:	gap of 100 bp	in length
*	22177	23040:	contlg of 864 bp	in length
*	23041	23140:	gap of 100 bp	in length
*	23141	24000:	contlg of 860 bp	in length
*	24001	24100:	gap of 100 bp	in length
*	24101	24966:	contlg of 866 bp	in length
*	24967	25066:	gap of 100 bp	in length
*	25067	25954:	contlg of 888 bp	in length
*	25955	26054:	gap of 100 bp	in length
*	26055	26916:	contlg of 862 bp	in length
*	26917	27016:	gap of 100 bp	in length
*	27017	27891:	contlg of 875 bp	in length
*	27892	27991:	gap of 100 bp	in length
*	27992	28859:	contlg of 868 bp	in length
*	28860	28959:	gap of 100 bp	in length
*	28960	29827:	contlg of 868 bp	in length
*	29828	29927:	gap of 100 bp	in length
*	29928	30828:	contlg of 901 bp	in length
*	30829	30928:	gap of 100 bp	in length
*	30929	31742:	contlg of 814 bp	in length
*	31743	31842:	gap of 100 bp	in length
*	31843	32699:	contlg of 857 bp	in length
*	32700	32799:	gap of 100 bp	in length
*	32800	33662:	contlg of 863 bp	in length
*	33663	33762:	gap of 100 bp	in length
*	33763	34638:	contlg of 876 bp	in length
*	34639	34738:	gap of 100 bp	in length
*	34739	35603:	contlg of 865 bp	in length
*	35604	35703:	gap of 100 bp	in length
*	35704	36578:	contlg of 875 bp	in length
*	36579	36678:	gap of 100 bp	in length
*	36679	37519:	contlg of 841 bp	in length

*	37520	37619:	gap of	100	bp	in	length
*	37620	38483:	contig of	864	bp	in	length
*	38484	38583:	gap of	100	bp	in	length
*	38584	39427:	contig of	844	bp	in	length
*	39428	39527:	gap of	100	bp	in	length
*	39528	40387:	contig of	860	bp	in	length
*	40388	40487:	gap of	100	bp	in	length
*	40488	41342:	contig of	855	bp	in	length
*	41343	41442:	gap of	100	bp	in	length
*	41443	42291:	contig of	849	bp	in	length
*	42292	42391:	gap of	100	bp	in	length
*	42392	43246:	contig of	855	bp	in	length
*	43247	43346:	gap of	100	bp	in	length
*	43347	44206:	contig of	860	bp	in	length
*	44207	44306:	gap of	100	bp	in	length
*	44307	45145:	contig of	843	bp	in	length
*	45150	45249:	gap of	100	bp	in	length
*	45250	46137:	contig of	888	bp	in	length
*	46138	46237:	gap of	100	bp	in	length
*	46238	47122:	contig of	885	bp	in	length
*	47123	47222:	gap of	100	bp	in	length
*	47223	48148:	contig of	926	bp	in	length
*	48149	48248:	gap of	100	bp	in	length
*	48249	49112:	contig of	864	bp	in	length
*	49113	49212:	gap of	100	bp	in	length
*	49213	50075:	contig of	863	bp	in	length
*	50076	50175:	gap of	100	bp	in	length
*	50176	51027:	contig of	852	bp	in	length
*	51028	51127:	gap of	100	bp	in	length
*	51128	51982:	contig of	855	bp	in	length
*	51983	52082:	gap of	100	bp	in	length
*	52083	52941:	contig of	859	bp	in	length
*	52942	53041:	gap of	100	bp	in	length
*	53042	53907:	contig of	866	bp	in	length
*	53908	54007:	gap of	100	bp	in	length
*	54008	54858:	contig of	851	bp	in	length
*	54859	54958:	gap of	100	bp	in	length
*	54959	55820:	contig of	862	bp	in	length
*	55821	55920:	gap of	100	bp	in	length
*	55921	56793:	contig of	873	bp	in	length
*	56794	56893:	gap of	100	bp	in	length
*	56894	57786:	contig of	893	bp	in	length
*	57787	57886:	gap of	100	bp	in	length
*	57887	58511:	contig of	625	bp	in	length
*	58512	58611:	gap of	100	bp	in	length
*	58612	59483:	contig of	872	bp	in	length
*	59484	59583:	gap of	100	bp	in	length
*	59584	60433:	contig of	856	bp	in	length
*	60440	60539:	gap of	100	bp	in	length
*	60540	61400:	contig of	861	bp	in	length
*	61401	61500:	gap of	100	bp	in	length
*	61501	62357:	contig of	857	bp	in	length
*	62358	62457:	gap of	100	bp	in	length
*	62458	63322:	contig of	868	bp	in	length
*	63326	63425:	gap of	100	bp	in	length
*	63426	64283:	contig of	858	bp	in	length
*	64284	64383:	gap of	100	bp	in	length
*	64384	65260:	contig of	877	bp	in	length
*	65261	65360:	gap of	100	bp	in	length
*	65361	66214:	contig of	854	bp	in	length
*	66215	66314:	gap of	100	bp	in	length
*	66315	67195:	contig of	881	bp	in	length
*	67196	67295:	gap of	100	bp	in	length
*	67296	68164:	contig of	889	bp	in	length
*	68185	68284:	gap of	100	bp	in	length
*	68285	69165:	contig of	898	bp	in	length
*	69183	69282:	gap of				

[illegible]

RESULT	6
AC069214/c	
LOCUS	AC069214 244410 bp DNA
DEFINITION	Homo sapiens chromosome 12 clone RP11-665J20, SEQUENCE, 28 unordered pieces.
ACCESSION	AC069214
VERSION	AC069214.5 GI:9438534
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 24410)

Muzny, D.M., Adams, C., Bailey, M., Barberia, J., Blankenburg, K.,
AUTHORS

Boboia, B., Bouck, J., Bowie, S., Brooks, A., Buhey, C., Bunea, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Dublin, K.J., Fernandez, C., Ferraguto, D., Fotum-Taisey, J., Frantz, P., Gansh, R., Gorrell, J.H., Gorrell, L., Guaveana, W., Harris, K., Hernandez, J., Hodgson, A., Hognes, M., Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, K., Kelly, S., Komdejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Locardo, R., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Neilson, A., Neuyen, R., Nguyen, N., Nguyen, S., Oswal, G., Patish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Oulles, M., Reiter, D., Rives, M., Sammel, S., Say, J., Scherer, S., Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, K., Taboor, P., Taylor, T., Vasquez, L., Vilson, R., Vo, O., Wabadi, M., Wallington, S., Welnsstock, G., Welnsstock, I.R., Williamson, A., Wolley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Neilson, D. and Gibbs, R.

TITLE	Direct Submission
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 244410)
AUTHORS	Morley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jul 25, 2000 this sequence version replaced gi:8468712.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAVI
Center clone name: RP11-665J20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208190 bases at least Q40
Consensus quality: 224572 bases at least Q30
Consensus quality: 231987 bases at least Q20
Estimated insert size: 219523; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 5.8x in Q20 bases; sum-of-contigs estimation

```

- * NOTE: Estimated insert size may differ from sequence length
- * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.htm)
- * NOTE: This is a 'working draft' sequence. It currently

QY 14 cggccagcgcccccggcgcccgagagccccaagctgccaagccggccccaagagtcg 73


```
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1      802: contig of 802 bp in length
* 803 902: gap of 100 bp
* 903 1682: contig of 780 bp in length
* 1683 1782: gap of 100 bp
* 1783 2579: contig of 797 bp in length
* 2580 2679: gap of 100 bp
* 2680 3486: contig of 807 bp in length
* 3487 3586: gap of 100 bp
* 3587 4397: contig of 811 bp in length
* 4398 4497: gap of 100 bp
* 4498 5301: contig of 804 bp in length
* 5302 5401: gap of 100 bp
* 5402 6192: contig of 791 bp in length
* 6193 6292: gap of 100 bp
* 6293 7085: contig of 793 bp in length
* 7086 7185: gap of 100 bp
* 7186 7913: contig of 728 bp in length
* 7914 8013: gap of 100 bp
* 8014 8807: contig of 794 bp in length
* 8808 8907: gap of 100 bp
* 8908 9695: contig of 788 bp in length
* 9696 9795: gap of 100 bp
* 9796 10586: contig of 791 bp in length
* 10587 10686: gap of 100 bp
* 10687 11492: contig of 806 bp in length
* 11493 11592: gap of 100 bp
* 11593 12284: contig of 692 bp in length
* 12285 12384: gap of 100 bp
* 12385 13191: contig of 807 bp in length
* 13192 13291: gap of 100 bp
* 13292 14071: contig of 780 bp in length
* 14072 14171: gap of 100 bp
* 14172 14981: contig of 810 bp in length
* 14982 15081: gap of 100 bp
* 15082 15904: contig of 823 bp in length
* 15905 16004: gap of 100 bp
* 16005 16776: contig of 772 bp in length
* 16777 16876: gap of 100 bp
* 16877 17672: contig of 796 bp in length
* 17673 17772: gap of 100 bp
* 17773 18553: contig of 781 bp in length
* 18554 18653: gap of 100 bp
* 18654 19448: contig of 795 bp in length
* 19449 19548: gap of 100 bp
* 19549 20346: contig of 798 bp in length
* 20347 20446: gap of 100 bp
* 20447 21243: contig of 797 bp in length
* 21244 21343: gap of 100 bp
* 21344 22130: contig of 787 bp in length
* 22131 22230: gap of 100 bp
* 22231 23021: contig of 791 bp in length
* 23022 23121: gap of 100 bp
* 23122 23926: contig of 805 bp in length
* 23927 24026: gap of 100 bp
* 24027 24818: contig of 792 bp in length
* 24819 24918: gap of 100 bp
* 24919 25648: contig of 730 bp in length
* 25649 25748: gap of 100 bp
* 25749 26549: contig of 801 bp in length
* 26550 26649: gap of 100 bp
* 26650 27407: contig of 758 bp in length
* 27408 27507: gap of 100 bp
* 27508 28281: contig of 774 bp in length
* 28282 28381: gap of 100 bp
* 28382 29164: contig of 783 bp in length
* 29165 29264: gap of 100 bp
*
* 29265 29979: contig of 715 bp in length
* 29980 30079: gap of 100 bp
* 30080 30860: contig of 781 bp in length
* 30861 30960: gap of 100 bp
* 30961 31755: contig of 795 bp in length
* 31756 31855: gap of 100 bp
* 31856 32630: contig of 775 bp in length
* 32631 32730: gap of 100 bp
* 32731 33532: contig of 802 bp in length
* 33533 33632: gap of 100 bp
* 33633 34348: contig of 716 bp in length
* 34349 34448: gap of 100 bp
* 34449 35207: contig of 759 bp in length
* 35208 35307: gap of 100 bp
* 35308 36102: contig of 795 bp in length
* 36103 36202: gap of 100 bp
* 36203 37012: contig of 810 bp in length
* 37013 37112: gap of 100 bp
* 37113 37889: contig of 777 bp in length
* 37890 37989: gap of 100 bp
* 37990 38772: contig of 783 bp in length
* 38773 38872: gap of 100 bp
* 38873 39615: contig of 743 bp in length
* 39616 39715: gap of 100 bp
* 39716 40522: contig of 807 bp in length
* 40523 40622: gap of 100 bp
* 40623 41422: contig of 800 bp in length
* 41423 41522: gap of 100 bp
* 41523 42327: contig of 805 bp in length
* 42328 42427: gap of 100 bp
* 42428 43250: contig of 823 bp in length
* 43251 43350: gap of 100 bp
* 43351 44028: contig of 678 bp in length
* 44029 44128: gap of 100 bp
* 44129 44925: contig of 797 bp in length
* 44926 45025: gap of 100 bp
* 45026 45830: contig of 805 bp in length
* 45831 45930: gap of 100 bp
* 45931 46699: contig of 769 bp in length
* 46700 46799: gap of 100 bp
* 46800 47580: contig of 781 bp in length
* 47581 47680: gap of 100 bp
* 47681 48469: contig of 789 bp in length
* 48470 48569: gap of 100 bp
* 48570 49348: contig of 779 bp in length
* 49349 49448: gap of 100 bp
* 49449 50261: contig of 813 bp in length
* 50262 50361: gap of 100 bp
* 50362 51166: contig of 805 bp in length
* 51167 51266: gap of 100 bp
* 51267 52069: contig of 803 bp in length
* 52070 52169: gap of 100 bp
* 52170 52968: contig of 799 bp in length
* 52969 53068: gap of 100 bp
* 53069 53858: contig of 790 bp in length
* 53859 53958: gap of 100 bp
* 53959 54711: contig of 753 bp in length
* 54712 54811: gap of 100 bp
* 54812 55489: contig of 678 bp in length
* 55490 55589: gap of 100 bp
* 55590 56181: contig of 552 bp in length
* 56182 56281: gap of 100 bp
* 56282 57080: contig of 799 bp in length
* 57081 57180: gap of 100 bp
* 57181 57920: contig of 740 bp in length
* 57921 58020: gap of 100 bp
* 58021 58733: contig of 713 bp in length
* 58734 58833: gap of 100 bp
* 58834 59552: contig of 719 bp in length
* 59553 59652: gap of 100 bp
* 59653 60419: contig of 767 bp in length
* 60420 60519: gap of 100 bp
```


Query Match	29.5%	Score 52.2	DB 54	Length 131602
Best Local Similarity	56.8%	Pred No. 1.3		
Matches 96, Conservative	0	Mismatches 73	Indels 0	Gaps 0
QY 5	ccgaggaggtcgagcccgagcgcccgagcgcccgagcgagcccgagcgcccgagcgcccg	64		
DB 83558	ccgg	83799		
QY 65	aggagtcggcgagcgagcccgagcgagcgagcgagcgagcgagcgagcgagcgagcgag	124		
DB 83798	cc	83739		
QY 125	caggggcccgc	173		
DB 83738	cgcc	83690		

RESULT	9
AC016541/c	
LOCUS	
DEFINITION	Homo sapiens clone RP11-20M3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC016541.2
VERSION	GI:9123972
KEYWORDS	HTG; HTGS; PHASED.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryote: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 78690)
TITLE	Barren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Homo sapiens, clone RP11-20M3
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 78690)
TITLE	Barren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Brown,J., Castle,A., Colangelo,M., Collins,S., Colliore,A., Cooke,P., DeRellano,K., Dewar,K., Domino,M., Doneanu,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kamm,L., Karatas,A., Klein,J., Lemocky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Teste,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL	Direct Submission
COMMENT	Submitted (30-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Jul 13, 2000 this sequence version replaced gi:6513976. All repeats were identified using RepeatMasker: Smt, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
	----- Genome Center
	Center: Whitehead Institute/ MIT Center for Genome Research
	Center code: WtBR
	Web site: http://www-seq.wi.mit.edu
	Contact: sequence_submissions@genome.wi.mit.edu
	----- Project Information
	Center project name: L3938
	Center clone name: 20_M_3

* NOTE: This record contains 81 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that

* the record is updated, the accession number will
* be preserved.
1
* 839 938: contig of 838 bp in length
* 939 1827: contig of 889 bp in length
* 1828 1927: gap of 100 bp
* 1928 2811: contig of 884 bp in length
* 2812 2911: gap of 100 bp
* 2912 3744: contig of 833 bp in length
* 3745 3844: gap of 100 bp
* 3845 4762: contig of 918 bp in length
* 4763 4862: gap of 100 bp
* 4863 5733: contig of 871 bp in length
* 5734 5833: gap of 100 bp
* 5834 6688: contig of 865 bp in length
* 6689 6798: gap of 100 bp
* 6799 7671: contig of 873 bp in length
* 7672 7771: gap of 100 bp
* 7772 8633: contig of 862 bp in length
* 8634 8733: gap of 100 bp
* 8734 9578: contig of 845 bp in length
* 9579 9678: gap of 100 bp
* 9679 10560: contig of 882 bp in length
* 10561 10660: gap of 100 bp
* 10661 11550: contig of 890 bp in length
* 11551 11650: gap of 100 bp
* 11651 12510: contig of 860 bp in length
* 12511 12610: gap of 100 bp
* 12611 13446: contig of 836 bp in length
* 13447 13546: gap of 100 bp
* 13547 14426: contig of 880 bp in length
* 14427 14526: gap of 100 bp
* 14527 15417: contig of 891 bp in length
* 15418 15517: gap of 100 bp
* 15518 16378: contig of 861 bp in length
* 16379 16478: gap of 100 bp
* 16479 17337: contig of 859 bp in length
* 17338 17437: gap of 100 bp
* 17438 18328: contig of 891 bp in length
* 18329 18428: gap of 100 bp
* 18429 19298: contig of 870 bp in length
* 19299 19398: gap of 100 bp
* 19399 20220: contig of 822 bp in length
* 20221 20320: gap of 100 bp
* 20321 21214: contig of 894 bp in length
* 21215 21314: gap of 100 bp
* 21315 22202: contig of 888 bp in length
* 22203 22302: gap of 100 bp
* 22303 23160: contig of 858 bp in length
* 23161 23260: gap of 100 bp
* 23261 24084: contig of 824 bp in length
* 24085 24184: gap of 100 bp
* 24185 25053: contig of 869 bp in length
* 25054 25153: gap of 100 bp
* 25154 26032: contig of 879 bp in length
* 26033 26132: gap of 100 bp
* 26133 27010: contig of 878 bp in length
* 27011 27110: gap of 100 bp
* 27111 28001: contig of 891 bp in length
* 28002 28101: gap of 100 bp
* 28102 28999: contig of 898 bp in length
* 29000 29099: gap of 100 bp
* 29100 29958: contig of 859 bp in length
* 29959 30058: gap of 100 bp
* 30059 30948: contig of 890 bp in length
* 30949 31048: gap of 100 bp
* 31049 31936: contig of 888 bp in length
* 31937 32036: gap of 100 bp
* 32037 32919: contig of 883 bp in length
* 32920 33019: gap of 100 bp
* 33020 33869: contig of 850 bp in length
* 33870 33969: gap of 100 bp
* 33970 34850: contig of 881 bp in length

* 34851 34950: gap of 100 bp
* 34951 35844: contig of 894 bp in length
* 35845 35944: gap of 100 bp
* 35945 36851: contig of 907 bp in length
* 36852 36951: gap of 100 bp
* 36952 37848: contig of 897 bp in length
* 37849 37948: gap of 100 bp
* 37949 38744: contig of 796 bp in length
* 38745 38844: gap of 100 bp
* 38845 39745: contig of 901 bp in length
* 39746 39845: gap of 100 bp
* 39846 40711: contig of 866 bp in length
* 40712 40811: gap of 100 bp
* 40812 41688: contig of 877 bp in length
* 41689 41788: gap of 100 bp
* 41789 42644: contig of 856 bp in length
* 42645 42744: gap of 100 bp
* 42745 43612: contig of 868 bp in length
* 43613 43712: gap of 100 bp
* 43713 44559: contig of 847 bp in length
* 44560 44659: gap of 100 bp
* 44660 45529: contig of 870 bp in length
* 45530 45629: gap of 100 bp
* 45630 46515: contig of 886 bp in length
* 46516 46615: gap of 100 bp
* 46616 47501: contig of 886 bp in length
* 47502 47601: gap of 100 bp
* 47602 48484: contig of 883 bp in length
* 48485 48584: gap of 100 bp
* 48585 49438: contig of 854 bp in length
* 49439 49538: gap of 100 bp
* 49539 50413: contig of 875 bp in length
* 50414 50513: gap of 100 bp
* 50514 51390: contig of 877 bp in length
* 51391 51490: gap of 100 bp
* 51491 52359: contig of 869 bp in length
* 52360 52459: gap of 100 bp
* 52460 53320: contig of 861 bp in length
* 53321 53420: gap of 100 bp
* 53421 54316: contig of 896 bp in length
* 54317 54416: gap of 100 bp
* 54417 55274: contig of 858 bp in length
* 55275 55374: gap of 100 bp
* 55375 56245: contig of 871 bp in length
* 56246 56345: gap of 100 bp
* 56346 57209: contig of 864 bp in length
* 57210 57309: gap of 100 bp
* 57310 58212: contig of 903 bp in length
* 58213 58312: gap of 100 bp
* 58313 59192: contig of 880 bp in length
* 59193 59292: gap of 100 bp
* 59293 60169: contig of 877 bp in length
* 60170 60269: gap of 100 bp
* 60270 61097: contig of 828 bp in length
* 61098 61197: gap of 100 bp
* 61198 62100: contig of 903 bp in length
* 62101 62200: gap of 100 bp
* 62201 63095: contig of 895 bp in length
* 63096 63195: gap of 100 bp
* 63196 64079: contig of 884 bp in length
* 64080 64179: gap of 100 bp
* 64180 65024: contig of 845 bp in length
* 65025 65124: gap of 100 bp
* 65125 65999: contig of 875 bp in length
* 66000 66099: gap of 100 bp
* 66100 66973: contig of 874 bp in length
* 66974 67073: gap of 100 bp
* 67074 67946: contig of 873 bp in length
* 67947 68046: gap of 100 bp
* 68047 68942: contig of 896 bp in length
* 68943 69042: gap of 100 bp
* 69043 69930: contig of 888 bp in length
* 69931 70030: gap of 100 bp

[illegible]

[illegible][illegible]

[illegible]

1594 1693: gap of 100 bp
1694 2467: contig of 774 bp in length
2468 2567: gap of 100 bp
2568 3331: contig of 764 bp in length
3332 3431: gap of 100 bp
3432 4207: contig of 776 bp in length
4208 4307: gap of 100 bp
4308 5069: contig of 762 bp in length
5070 5169: gap of 100 bp
5170 5929: contig of 760 bp in length
5930 6029: gap of 100 bp
6030 6816: contig of 787 bp in length
6817 6916: gap of 100 bp
6917 7666: contig of 750 bp in length
7667 7766: gap of 100 bp
7767 8540: contig of 774 bp in length
8541 8640: gap of 100 bp
8641 9401: contig of 761 bp in length
9402 9501: gap of 100 bp
9502 10274: contig of 773 bp in length
10275 10374: gap of 100 bp
10375 11122: contig of 748 bp in length
11123 11222: gap of 100 bp
11223 11969: contig of 747 bp in length
11970 12069: gap of 100 bp
12070 12830: contig of 761 bp in length
12831 12930: gap of 100 bp
12931 13694: contig of 764 bp in length
13695 13794: gap of 100 bp
13795 14568: contig of 774 bp in length
14569 14668: gap of 100 bp
14669 15422: contig of 754 bp in length
15423 15522: gap of 100 bp
15523 16276: contig of 754 bp in length
16277 16376: gap of 100 bp
16377 17095: contig of 719 bp in length
17096 17195: gap of 100 bp
17196 17954: contig of 759 bp in length
17955 18054: gap of 100 bp
18055 18826: contig of 772 bp in length
18827 18926: gap of 100 bp
18927 19689: contig of 763 bp in length
19690 19789: gap of 100 bp
19790 20568: contig of 779 bp in length
20569 20668: gap of 100 bp
20669 21442: contig of 774 bp in length
21443 21542: gap of 100 bp
21543 22305: contig of 763 bp in length
22306 22405: gap of 100 bp
22406 23262: contig of 757 bp in length
23263 23262: gap of 100 bp
23263 24040: contig of 778 bp in length
24041 24140: gap of 100 bp
24141 24889: contig of 749 bp in length
24890 24989: gap of 100 bp
24990 25763: contig of 774 bp in length
25764 25863: gap of 100 bp
25864 26646: contig of 783 bp in length
26647 26746: gap of 100 bp
26747 27508: contig of 762 bp in length
27509 27608: gap of 100 bp
27609 28367: contig of 759 bp in length
28368 28467: gap of 100 bp
28468 29245: contig of 778 bp in length
29246 29345: gap of 100 bp
29346 30114: contig of 769 bp in length
30115 30214: gap of 100 bp
30215 30970: contig of 756 bp in length
30971 31070: gap of 100 bp
31071 31839: contig of 769 bp in length
31840 31939: gap of 100 bp
31940 32654: contig of 715 bp in length
32655 32754: gap of 100 bp

32755 33517: contig of 763 bp in length
33518 33617: gap of 100 bp
33618 34378: contig of 762 bp in length
34380 34479: gap of 100 bp
34480 35241: contig of 762 bp in length
35242 35341: gap of 100 bp
35342 36077: contig of 736 bp in length
36078 36177: gap of 100 bp
36178 36927: contig of 750 bp in length
36928 37027: gap of 100 bp
37028 37777: contig of 750 bp in length
37778 37877: gap of 100 bp
37878 38623: contig of 746 bp in length
38624 38723: gap of 100 bp
38724 38493: contig of 770 bp in length
39494 39593: gap of 100 bp
39594 40360: contig of 767 bp in length
40361 40460: gap of 100 bp
40461 41219: contig of 759 bp in length
41220 41319: gap of 100 bp
41320 42090: contig of 771 bp in length
42091 42190: gap of 100 bp
42191 42953: contig of 763 bp in length
42954 43053: gap of 100 bp
43054 43829: contig of 776 bp in length
43830 43929: gap of 100 bp
43930 44697: contig of 768 bp in length
44698 44797: gap of 100 bp
44798 45557: contig of 760 bp in length
45558 45657: gap of 100 bp
45658 46395: contig of 738 bp in length
46396 46495: gap of 100 bp
46496 47248: contig of 753 bp in length
47249 47348: gap of 100 bp
47349 48118: contig of 770 bp in length
48119 48218: gap of 100 bp
48219 48976: contig of 758 bp in length
48977 49076: gap of 100 bp
49077 49834: contig of 758 bp in length
49835 49934: gap of 100 bp
49935 50692: contig of 758 bp in length
50693 50792: gap of 100 bp
50793 51517: contig of 725 bp in length
51518 51617: gap of 100 bp
51618 52351: contig of 734 bp in length
52352 52451: gap of 100 bp
52452 53215: contig of 764 bp in length
53216 53315: gap of 100 bp
53316 53968: contig of 654 bp in length
53970 54069: gap of 100 bp
54070 54828: contig of 759 bp in length
54829 54928: gap of 100 bp
54929 55690: contig of 762 bp in length
55691 55790: gap of 100 bp
55791 56554: contig of 764 bp in length
56555 56654: gap of 100 bp
56655 57423: contig of 769 bp in length
57424 57523: gap of 100 bp
57524 58295: contig of 772 bp in length
58296 58395: gap of 100 bp
58396 59176: contig of 781 bp in length
59177 59276: gap of 100 bp

Query Match 27.5%; Score 48.6; DB 53; Length 68581;
Best Local Similarity 53.8%; Pred. No. 6.2;
Matches 93; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 5 ccgaggggtcgcccgagcgcccgccgagggagagcccgagcgcccgcc 64
Db 38437 CCAAGGCGCAACCAAGGGGCGCCCGCCGACGCGCCCGCGCGCCG 38378
QY 65 aggagtcgcgcgcgagcgcccgcccgcggtacgtcttgaacgcgagcgagcgccgc 124


```

22 116.2 26.3 1554 37 AF269249 Homo sapi
23 115.2 26.1 2330 37 AK023989 Homo sapi
24 115.2 26.1 3000 37 AF060503 Homo sapi
25 115 26.1 1909 37 AF242768 Homo sapi
26 114.4 25.9 166803 55 AC048382 Homo sapi
27 114.4 25.9 194486 41 AC016771 Homo sapi
28 113.8 25.8 2454 11 AB041616 Homo sapi
29 113.8 25.8 259451 40 AC015605 Mus muscu
30 113.6 25.8 183525 39 AC012313 Homo sapi
31 111.4 25.3 2692 33 AK024191 Homo sapi
32 111.4 25.3 124237 66 AF189001 Homo sapi
33 110.4 25.0 124237 66 AF189001 Homo sapi
34 109.6 24.9 259451 40 AC015605 Mus muscu
35 108.8 24.7 2587 10 AF055077 Homo sapi
36 108.8 24.7 2678 85 HUM2E1 Homo sapi
37 108.8 24.7 2905 37 AF055078 Homo sapi
38 108.8 24.7 12456 37 AF161886 Homo sapi
39 108.8 24.7 40616 41 AC016630 Homo sapi
40 108.8 24.7 44210 41 AC016629 Homo sapi
41 108.4 24.6 4582 11 M05PLM24X Mus muscu
42 108.4 24.6 203100 55 AC060772 Mus muscu
43 108.2 24.5 113247 65 AC079506 Mus muscu
44 108.2 24.5 168700 30 AC009115 Homo sapi
45 108.2 24.5 230534 65 AC079575 Mus muscu

```

ALIGNMENTS

```

RESULT 1
LOCUS AK025356 2878 bp mRNA PRI
DEFINITION Homo sapiens CDNA: FLJ21703 fls, clone COL09895, highly similar to
HS245553 Homo sapiens mRNA for AP4 zinc finger protein.
ACCESSION AK025356
VERSION AK025356.1 GI:10437856
KEYWORDS Oligo capping; fls (full insert sequence).
SOURCE Homo sapiens colon CDNA to: mRNA, clone.lib:COL clone:COL09895.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEO human cDNA sequencing project
JOURNAL Unpublished (2000)
AUTHORS 2 (bases 1 to 2878)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center,
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; CDNA full insert
sequencing: Research Association for Biotechnology; CDNA library
construction: 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
SOURCE 1..2878
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL09895"
/clone.lib="COL"
/tissue.type="Colon"
/note="cloning vector pME18SFL3"
misc_feature 1..2878

```

```

BASE COUNT 560 a 974 c 826 g 518 t
ORIGIN

```

```

Query Match 100.0%; Score 441; DB 37; Length 2878;
Best Local Similarity 100.0%; Pred. No. 5e-62;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggtgagatgacgctgacgacccctccacgtgtgctgtgtggcaagcgttcggac 60
DB 706 GGTGAGATGACGCTGACGACCCCTCCACGTGTGCTGTGTGGCAAGCGCTTCGGCAC
QY 61 aagcccaattatgctacacgacgacgacgacgacgacgacgacgacgacgacgac 120
DB 766 AAGCCCAATTATGCTACACGACGACGACGACGACGACGACGACGACGACGACGAC
QY 121 ggtgagatgacgctgacgacccctccacgtgtgctgtgtggcaagcgttcggac 180
DB 826 GGTGAGATGACGCTGACGACCCCTCCACGTGTGCTGTGTGGCAAGCGCTTCGGCAC
QY 181 ggtgagatgacgctgacgacccctccacgtgtgctgtgtggcaagcgttcggac 240
DB 886 GGTGAGATGACGCTGACGACCCCTCCACGTGTGCTGTGTGGCAAGCGCTTCGGCAC
QY 241 ggtgagatgacgctgacgacccctccacgtgtgctgtgtggcaagcgttcggac 300
DB 946 GGTGAGATGACGCTGACGACCCCTCCACGTGTGCTGTGTGGCAAGCGCTTCGGCAC
QY 301 ggtgagatgacgctgacgacccctccacgtgtgctgtgtggcaagcgttcggac 360
DB 1006 GGTGAGATGACGCTGACGACCCCTCCACGTGTGCTGTGTGGCAAGCGCTTCGGCAC
QY 361 ggtgagatgacgctgacgacccctccacgtgtgctgtgtggcaagcgttcggac 420
DB 1066 GGTGAGATGACGCTGACGACCCCTCCACGTGTGCTGTGTGGCAAGCGCTTCGGCAC
QY 421 ggtgagatgacgctgacgacccctccacgtgtgctgtgtggcaagcgttcggac 480
DB 1126 GGTGAGATGACGCTGACGACCCCTCCACGTGTGCTGTGTGGCAAGCGCTTCGGCAC

```

```

RESULT 2
LOCUS HSA245553 2878 bp mRNA PRI
DEFINITION Homo sapiens mRNA for AP4 zinc finger protein.
ACCESSION AJ245553
VERSION AJ245553.1 GI:5748564
KEYWORDS AP4 gene; AP4 protein; zinc finger protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2878)
AUTHORS Dobner,T.G., Fischer,M. and Groft,P.
TITLE Cloning of a novel zinc finger protein
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2878)
Dobner,T.G.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1999) Dobner T.G., Institut fuer Medizinische
Mikrobiologie und Hygiene, Universitaet Regensburg,
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, GERMANY
FEATURES
SOURCE 1..2878
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Hela"
106..1776
/gene="AP4"
106..1776
/gene="AP4"

```



```

/note="zinc finger protein"
/codon_start=1
/product="Ap4 protein"
/protein_id="CAB53100.1"
/db_xref="GI:5748365"
/translation="MGAPQRLSLGPSQSPQTLGKESRGLRQOGTSVAQSGAQBGR
AHACACRRHFPQVALMLHTRCOARLPDPECCRRFMRKOLRAHLOHCHPDA
LGFACHLCOSEFGWVALVHLRAHSAKRPICPRCERFRMRKOLRAHLOHCHPDA
PAARPTCCNGCSFQMDQVLAHNRVHALEAALAAKVLGSRSGRAVTPAPRG
DAVDRPQCACCKRRFNRHNLIAHRRVHGERPHOCPEGCKRFTNKPYLTSRRLHT
GKRPYCKECGRFRFRHKNLISHKILHKSSEGAQAAPGSPQLAPGQESAEPPLP
ADPLKPAOEPPGAPPEHPQDPIEAPPSLYSCDDCRSEFLERFLRAHROHGEHPF
TCAECNKGKETHLVASHSVHSGERFACCEGSRFSGSHLAHRDHAADPRPYC
PDGKAFHKKPYLAHRRIHTGKPYVCDGKAFKSNLVSHRRIHTGERSYACPD
CDRSEFSKSNLHTRKSHIRIDGAFCCALICGQTFDDEERLLAHOKKHDV"

BASE COUNT      528 a      973 c      839 g      538 t
ORIGIN

Query Match      100.0%; Score 441; DB 77; Length 2878;
Best Local Similarity 100.0%; Pred. No. 5e-62;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtgagatgcctgcagccgcccccttcagtgctgctgtgtgcaagcgcttcgagc 60
|||||
DB 754 GGTGAGATGCCCTGCAGCCGCCCTTCAGTGTGCTGTTGTGGCAAGCGCTTCGGGAC 813

QY 61 aagcccaattgatgcctacccgcgctgcacagggcgagcgcccccaagtgcccc 120
|||||
DB 814 AAGCCCAACTGATGCTCACCGCGGTGCACAGGGCGCGCCGCCCAACAGTGGCCCC 873

QY 121 gagtgcggagagcgcttaccataagccctatctgaacttcgacggcgatccacacc 180
|||||
DB 874 GAGTGCGGAGAGCGCTTACCAATTAAGCCCTATCTGACTGTGCACGGCGCATCAACAC 933

QY 181 ggcgagaagccctacccgttcgaaagagtgccgcgcgcttcgcgcgcaaacaccactg 240
|||||
DB 934 GCGGAGAGAGCGCTTACCGTGCAGAGAGTGGCGCGCGCTTCGGGCAACCAACCACTG 993

QY 241 ctgtctacagcaagattcaacaagagatccgagggatcgccgacggccccgcgcgcg 300
|||||
DB 994 CTGTCTACAGCAAGATTCAACAAGCATTCGAGGGGTGCGGCCAGGCGCGCCCGCGGCGG 1053

QY 301 gggagcccccaagctgcagcgccgcccagagatcgccgagccgccccgcgcgta 360
|||||
DB 1054 GGGAGCCCCCAGCTGCGAGCGCGGCCAGAGATCGCGCGCGGACCCCGCGGTGA 1113

QY 361 cctctgaaccgagcccgagagcgccgcgagggcgcccgccagagaccccgcaagaccg 420
|||||
DB 1114 CCTCTGAACCGCGGCCAGAGAGCGCGGCCCGCCGCAAGAGCAACCCGCAAGACCGC 1173

QY 421 atcgaagccccccctccctc 441
|||||
DB 1174 ATCGAAGCCCCCCCCCTCCCTC 1194

RESULT 3
AF201303      2954 bp      mRNA      PRI      20-JAN-2000
DEFINITION   Homo sapiens dhfr oribeta-binding protein RIP60 mRNA, complete cds.
ACCESSION   AF201303
VERSION      AF201303.1 GI:6716713
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 2954)
              Houchens,C.R., Montigny,W., Zeltser,L., Dailey,L., Gilbert,J.M. and
              Heintz,N.H.
              The dhfr oribeta-binding protein RIP60 contains 15 zinc fingers:
              DNA binding and looping by the central three fingers and an
              associated proline-rich region

```

```

JOURNAL      Nucleic Acids Res. 28 (2), 570-581 (2000)
PUBMED      1060657
REFERENCE    2 (bases 1 to 2954)
AUTHORS     Houchens,C.R., Gilbert,J.M. and Heintz,N.H.
TITLES      Direct Submission
JOURNAL     Submitted (02-NOV-1999) Pathology, University of Vermont, Soule
              Medical Alumni Building, Burlington, VT 05405, USA
              Location/Qualifiers
FEATURES
SOURCE
CDS
1..2954
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Hela"
130..1833
/note="contains 15 C2H2-Kruppel-like zinc fingers"
/codon_start=1
/product="dhfr oribeta-binding protein RIP60"
/protein_id="AF26712.1"
/db_xref="GI:6716714"
/translation="MLERRCRGLPLAMGLAQPRLLSGPSQSPQTLGKESRGLRQOGTS
VAQSGAQBGRHRAHCAHRRHPPGVALVHLHRAHSAKOPDPECCRRFMRKOL
HROYHAATMDLGFACHLCOSEFGWVALVHLRAHSAKOPDPECCRRFMRKOL
RAHLCRCHPAPAPARPTGNGCSRFAQMDQVLAHNRVHALEAALAAKVLGSR
RPAVTPAPRGDAVDRPQCACCKRRFNRHNLIAHRRVHGERPHOCPEGCKRFTNK
PYLTSRRLHTGKRPYCKECGRFRFRHKNLISHKILHKSSEGAQAAPGSPQLA
PQESAEPPLPADPLKPAOEPPGAPPEHPQDPIEAPPSLYSCDDCRSEFLERFL
HROHCHPDAARPTCCNGCSFQMDQVLAHNRVHSGERFACCEGSRFSGSHLAHR
PDHAPDRPYCDGKAFHKKPYLAHRRIHTGKPYVCDGKAFKSNLVSHRRIHT
HTRERPAKPCDRSEFSKSNLHTRKSHIRIDGAFCCALICGQTFDDEERLLAHOKK
H"

BASE COUNT      568 a      991 c      848 g      547 t
ORIGIN

Query Match      100.0%; Score 441; DB 37; Length 2954;
Best Local Similarity 100.0%; Pred. No. 4.9e-62;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtgagatgcctgcagccgcccccttcagtgctgctgtgtgcaagcgcttcgagc 60
|||||
DB 811 GGTGAGATGCCCTGCAGCCGCCCTTCAGTGTGCTGTTGTGGCAACCGCTTCGGGAC 870

QY 61 aagcccaattgatgcctacccgcgctgcacagggcgagcgcccccaagtgcccc 120
|||||
DB 871 AAGCCCAACTGATGCTCACCGCGGTGCACAGGGCGCGCCGCCCAACCACTG 930

QY 121 gagtgcggagagcgcttaccataagccctatctgaacttcgacggcgatccacacc 180
|||||
DB 931 GAGTGCGGAGAGCGCTTACCAATTAAGCCCTATCTGACTGTGCACGGCGCATCAACAC 990

QY 181 ggcgagaagccctacccgttcgaaagagtgccgcgcgcttcgcgcgcaaacaccactg 240
|||||
DB 991 GCGGAGAGAGCGCTTACCGTGCAGAGAGTGGCGCGCGCTTCGCGCAACCAACCACTG 1050

QY 241 ctgtctacagcaagattcaacaagagatccgagggatcgccgagccgccccgcgcgcg 300
|||||
DB 1051 CTGTCTACAGCAAGATTCAACAAGCATTCGAGGGGTGCGGCCAGGCGCGCCCGCGGTGA 1110

QY 301 gggagcccccaagctgcagcgccgcccagagatcgccgagccgccccgcgcgta 360
|||||
DB 1111 GGGAGCCCCCAGCTGCGAGCGCGGCCAGAGATCGCGCGCGGACCCCGCGGTGA 1170

QY 361 cctctgaaccgagcccgagagcgccgcgagggcgcccgccagagaccccgcaagaccg 420
|||||
DB 1171 CCTCTGAACCGCGGCCAGAGAGCGCGGCCCGCCGCAAGAGCAACCCGCAAGACCGC 1230

QY 421 atcgaagccccccctccctc 441
|||||
DB 1231 ATCGAAGCCCCCCCCCTCCCTC 1251

RESULT 4
AC005586/c    132150 bp    DNA      PRI      30-SEP-2000
LOCUS

```

DEFINITION	Homo sapiens PAC clone RP4-584D14 from Tg31-g35, complete sequence.
ACCESSION	AC005586
VERSION	AC005586.2
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 132150)
TITLE	Sulston, J.E. and Waterston, R.
JOURNAL	Toward a complete human genome sequence
MEDLINE	Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE	99063792
AUTHORS	2 (bases 1 to 132150)
TITLE	Stoneking, T., Ozersky, P., Woldmann, P. and Le, T.
JOURNAL	The sequence of Homo sapiens PAC clone RP4-584D14
REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 132150)
TITLE	Waterston, R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (01-SEP-1998) Genome Sequencing Center, Washington
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
TITLE	MO 63108, USA
JOURNAL	4. (bases 1 to 132150)
REFERENCE	Waterston, R.H.
AUTHORS	Direct Submission
TITLE	Submitted (12-JUN-2000) Genome Sequencing Center, Washington
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
REFERENCE	MO 63108, USA
AUTHORS	5 (bases 1 to 132150)
TITLE	Waterston, R.
JOURNAL	Direct Submission
REFERENCE	Submitted (30-SEP-2000) Department of Genetics, Washington
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE	On Jun 12, 2000 this sequence version replaced gi:3907529.
JOURNAL	
COMMENT	

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIGRI Chromosome 7 Mapping Project (Eric D. Green, director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/G7/CH7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RPc1-4, prepared by Pliered de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTORIZ: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP5-820A21, 200 base pair overlap. Actual start of this clone is at base position 1 of RP4-584BD14; actual end is at base position 131954 of RP4-584BD14.

The sequence RP4-584BD14 from base position 21655 to 22775 is GT rich. This region was sized with PCR from genomic DNA and the HindIII digest with band size 3643 real, and 3642 insilico. The sequence RP4-584BD14 from base position 22628 to 22637 is represented by sequence derived by PCR from genomic DNA.

Location/Qualifiers

1..132150

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31-q35"
/clone="RP4-584D14"
/clone_lib="RP4-4"
330.315
repeat_region
/rpt_family="MIR"
960.1288
repeat_region
/rpt_family="MER4-group"
1360.1468
repeat_region
/rpt_family="L2"
1562.1616
repeat_region
/rpt_family="L2"
1733.2080
misc_feature
/note="similar to EST AW510685 (NID:g7148763) hc89e08.x1"
2051.2114
repeat_region
/rpt_family="L2"
2148.3224
misc_feature
/note="CpG_island (GC=70.7, o/e=0.90, #CpGs=107)"
2347.2426
repeat_region
/rpt_family="GC-rich"
3141.3232
repeat_region
/rpt_family="Alu"
3235.3368
repeat_region
/rpt_family="Alu"
3795.3888
repeat_region
/rpt_family="MIR"
4037.4199
repeat_region
/rpt_family="Alu"
4308.4593
repeat_region
/rpt_family="MER4-group"
4872.5309
misc_feature
/note="similar to EST A1827463 (NID:g5448134) wf29f03.x1"
4882.4909
repeat_region
/rpt_family="AT-rich"
5679.5773
repeat_region
/rpt_family="MIR"
5947.5952
misc_feature
/note="match to EST AA482420 (NID:g2210098) zc34f02.r1"
6148.6450
repeat_region
/rpt_family="Alu"
6453.6690
repeat_region
/rpt_family="MER4-group"
6695.7007
repeat_region
/rpt_family="Alu"
7008.7117
repeat_region
/rpt_family="MER4-group"
7180.7355
repeat_region
/rpt_family="MER4-group"
7615
misc_feature
/note="match to EST A1144211 (NID:g3666020) qb93g11.x1"
7883.8397
misc_feature
/note="match to EST W72943 (NID:q1383235) zd54f12.r1"
7886.8357
misc_feature
/note="match to EST A1144211 (NID:g3666020) qb93g11.x1"
7886.8562
misc_feature

```

misc_feature	/note="similar to EST AW4111281 (NID:g6936822) fh1ld05.y1"
misc_feature	/note="similar to EST AW246540 (NID:g6589533) "
misc_feature	/note="match to EST AA129215 (NID:g1689084) zn84b02.s1"
STS	7886..8133 /db_xref="gi:1131717"
misc_feature	7886..8146 /note="similar to EST AI868891 (NID:g5542895) wc49f12.x1"
misc_feature	7888..8340 /note="match to EST H37790 (NID:g907289) yp46h05.s1"
misc_feature	7896..8372 /note="match to EST R72488 (NID:g846520) yj50g06.r1"
misc_feature	7976..8575 /note="match to EST R70860 (NID:g844377) y149h02.r1"
misc_feature	8008..8309 /note="match to EST AA335898 (NID:g1988160) "
misc_feature	8075..8451 /note="match to EST AA121027 (NID:g1678432) zn84g11.r1"
misc_feature	8192..8564 /note="similar to EST AM404837 (NID:g6923894) "
misc_feature	8228..8362 /note="match to EST AA040559 (NID:g1516964) zf06d08.r1"
misc_feature	8289..8732 /note="similar to EST AA843440 (NID:g292958) ak07h10.s1"
misc_feature	8302..8928 /note="match to EST AA059375 (NID:g1553139) zf66g07.r1"
misc_feature	8433..8746 /note="match to EST N45962 (NID:g1187128) yy34h02.r1"
misc_feature	8440..8968 /note="match to EST AA482420 (NID:g2210098) zt34f02.r1"
misc_feature	8459..8461 /note="match to EST AA481966 (NID:g2209644) zv43e06.r1"
misc_feature	8463..8928 /note="match to EST AA05336 (NID:g1543923) zf60d12.r1"
misc_feature	8591..9167 /note="match to EST AA129342 (NID:g1689125) zn84b02.r1"
misc_feature	8616..9159 /note="match to EST AA481966 (NID:g2209644) zv43e06.r1"
misc_feature	8707..8845 /note="similar to EST AW401523 (NID:g6920209) "
misc_feature	8743..8744 /note="match to EST R18324 (NID:g771934) yf59b11.r1"
misc_feature	8754..9069 /note="match to EST R18324 (NID:g771934) yf59b11.r1"
misc_feature	8776..9343 /note="match to EST N36800 (NID:g1157942) yy34h02.s1"
misc_feature	8826..9377 /note="similar to EST AI359215 (NID:g4110836) qy26g11.x1"
misc_feature	8835..9229 /note="similar to Oryctolagus cuniculus EST C83534 (NID:g3062491) "
misc_feature	8841..9377

Query Match	100.0%;	Score 441;	DB 9;	Length 132150;
Best Local Similarity	100.0%;	Pred. No. 9.5e-63;		
Matches 441;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	ggggaatgcccgtgagccgccccttcagaatgagcctgtgtggaaagcttcggagac	60
Db	10002	gctggaaatgcgcgcaccccccttccagatgctctgttgccaaaggcttccggcac	994
QY	61	aagcccaacttgatcgtctacacgcgcgctgtacacagcgagcgcgccacagtcgccc	120
Db	9942	aagcccaacttgaatgcgttcacgcgcgcgctgtacacagcgagcgcgccacagtcgccc	988
QY	121	gaatgctgggaagcgctttaccaaataagccctatctgacttgcaccgcgcatccaac	180
Db	9882	gagtgccgggaagcgctttaccaaataagccctatctgacttgcaccgcgcatccaac	982
QY	181	ggcgagaagcgcttaccgctgtgcaaaagtgtgcgcgcgcttcgagacaataccaaactg	240
Db	9822	ggcgagaagcgcttaccgctgtgcaaaagtgtgcgcgcgcttcgagacaataccaaactg	976

QY	241	ctgtctcaagcaagatctcaacagcgtatccgaagggtctgcgccaaagccgcctccgcgcgcg	300
Db	9762	CTGTCTCAGCAGCAGATTTCCAAAGCCATCCGAGGGGTGGCCAGGCCCTCCCTCCGGCCCG	9703
QY	301	ggagagcccccagctctgcagagccgcgcctcccaagatctccgcgcgcgcgcacccgcgcgcgt	360
Db	9702	GGGAGCCCCCAGCTGCCAGCCGGGCCCCCAAGAGTCCGGCGGCGAGCCACCCCGGGGTA	9643
QY	361	ccctctgaaacgc	420
Db	9642	CCCTTGAAACCGGCCCCGAGGCGCGCGCCGACGAGGCCCTCCAGAGCACCACGACGACCG	9583
QY	421	atcgaagcccccccctctccctc	441
Db	9582	ATCGAAGCCCCCTCTCTCTC	9562

RESULT	5
AC015887	
LOCUS	AC015887
DEFINITION	AC015887 201458 bp DNA HTG Mus musculus clone C77-3156, WORKING DRAFT SEQUENCE, 30-MAY-2000
ACCESSION	AC015887
VERSION	AC015887.3 GI:8112888
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	house mouse.

REFERENCE
AUTHORS
TITLE
Mus musculus chromosome, clone CT7-315E6
Birren, B., Linton, L., Nussbaum, C. and Lander, E.
1 (Passes 1 to 201458)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Mus musculus chromosome, clone CT7-315E6
 unpublished
 2 (bases 1 to 201458)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M

Authors
Baldwin, B., Linton, L., Nussbaum, C., Lande, E., Allen, N., Anderson, M., Baldwin, J., Barua, N., Becker, R., Boguslavsky, L., Bouknight, R., Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArliano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, S., Hagos, B., Hearford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kamm, L., Karatatz, A., Klein, J., Lenczky, J., Lien, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Strange-Shoman, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.-J., Zimmer, A. and Zody, M.

JOURNAL Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 30, 2000 this sequence version replaced gi:7249370.

----- Genome Center
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----

```

Center project name: L742
Center clone name: 315_E_6
----- Summary Statistics

```

Sequencing vector: M13; M77815; 448 of reads
Sequencing vector: Plasmid; n/a; 8-0.188 of reads
55.55555555556Chemistry: Dye-primer-amsrham; 588 of reads

Chemistry: Dye-terminator Big Dye; 428 of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 192611 bases at least Q40
Consensus quality: 196150 bases at least Q30
Consensus quality: 197810 bases at least Q20

[illegible][illegible]

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H.DJ0751H13

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-811N16, 200 bp overlap. Actual start of this clone is at base position 1 of RP4-751H13; actual end is at 128361 of RP4-751H13.

FEATURES

Source

repeat_region	/organism="Homo sapiens"	
repeat_region	/db_xref="taxon:9606"	
repeat_region	/chromosome="7"	
repeat_region	/map="7q35-qter"	
repeat_region	/clone="RP4-751H13"	
repeat_region	/clone_lib="unknown"	
repeat_region	20..155	
repeat_region	/rpt_family="Alu"	
repeat_region	860..941	
repeat_region	/rpt_family="MERL_type?"	
repeat_region	2351..2885	
repeat_region	/rpt_family="Retroviral"	
repeat_region	2924..3246	
repeat_region	/rpt_family="Alu"	
repeat_region	3300..3727	
repeat_region	/rpt_family="Retroviral"	
repeat_region	4146..4287	
repeat_region	/rpt_family="L2"	
repeat_region	5253..5728	
repeat_region	/rpt_family="MaLR"	
repeat_region	5933..6243	
repeat_region	/rpt_family="Alu"	
repeat_region	7787..7870	
repeat_region	/rpt_family="MIR"	
repeat_region	7798..8145	
repeat_region	/note="match to EST AA084720 (NID:g1626811) zn06f02.s1"	
repeat_region	8259..8311	
repeat_region	/rpt_family="L2"	
repeat_region	9100..9426	
repeat_region	/note="match to EST AA080867 (NID:g1623560) zn06f02.r1"	
repeat_region	9773..10086	
repeat_region	/rpt_family="Alu"	
repeat_region	/complement(9947..10370)	
repeat_region	/note="match to EST AA828832 (NID:g2901931) od80c03.s1"	
repeat_region	9947..10370	
repeat_region	/note="match to EST AA828832 (NID:g2901931) od80c03.s1"	
repeat_region	10066..10388	
repeat_region	/note="match to EST AA857387 (NID:g2945689) od83h10.s1"	
repeat_region	10297..10734	
repeat_region	/rpt_family="MaLR"	
repeat_region	11605..12220	
repeat_region	/rpt_family="L1"	
repeat_region	12222..12615	
repeat_region	/rpt_family="Retroviral"	
repeat_region	12617..12755	
repeat_region	/rpt_family="L1"	
repeat_region	12768..12890	

```

repeat_region      /rpt_family="Alu"
                   23679..23794
                   /rpt_family="MIR"
repeat_region      23773..23846
                   /rpt_family="L2"
repeat_region      24894..24982
                   /rpt_family="L2"
repeat_region      24983..25271
                   /rpt_family="Alu"
repeat_region      25272..25616
                   /rpt_family="L2"
misc_feature        26933..29260
note="Cpg_island (%GC=73.8, o/e=0.90, #CpGs=195)"

Query Match      28.7%; Score 126.6; DB 8; Length 128361;
Best Local Similarity 67.8%; Pred. No. 1.9e-12;
Matches 177; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 11 ccgttgaccgcgccctccagtgctgtgtgtggcaagcgtccggacaaagcccaact 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21081 CGGGGAGACGGGGCCCTACCCGTGCGGAGATGGGCAAGCGCTTCAGCAGAGATCCACC 21022

OY 71 tgatgcgtccacgcgcgtgtgacacagggagcagggcccccacagtgcccagtgcgga 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21021 TGGGCTCGCACCACAAAGACCCACACCGGAGAGCGGCCCTTCCCTTCACAGGATGGAGA 20962

OY 131 agcgcttaccataagccctatctgacttcgcacccgycgcatccacacccgycgaagaagc 190
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20961 AGCGCTTTCGCAAGAAAGACGCACTTGATTCGGCACACGCGCATCATACGGGCGAGAGAGC 20902

OY 191 cctaccgctgcaagaagtgtgcgcgcgcgttcctcgacaaacccaactgtctgtctaca 250
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20901 CCTACCACTGGCCACAGTGGCGGACGACGCTTCACGCACAGCAGCACTTGTCGGGACC 20842

OY 251 gcaagatccacaagcgatccg 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20841 AAGGCTGCACCGACGCGCG 20821

RESULT 9
AC013273/c
LOCUS AC013273 210923 bp DNA HTG 15-JUL-2000
DEFINITION Homo sapiens chromosome 7 clone RP11-428D5, WORKING DRAFT SEQUENCE.
ACCESSION AC013273
VERSION AC013273.5 GI:9211421
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mamalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 210923)
JOURNAL Waterston,R.H.
REFERENCE 2 (bases 1 to 210923)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 210923)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Genome Sequencing Center, Washington
AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis
JOURNAL MO 63108, USA
COMMENT On Jul 15, 2000 this sequence version replaced gi:899085.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0428D5
----- Summary Statistics -----
Sequencing vector: M13, 74%
Sequencing vector: plasmid, 26%
Chemistry: Dye-primed ET, 74% of reads

```

NOTE: THIS IS A WORKING DRAFT SEQUENCE. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	2429: contig of 2429 bp in length
*	2430	2529: gap of unknown length
*	2530	4617: contig of 2088 bp in length
*	4618	4717: gap of unknown length
*	4718	11807: contig of 7030 bp in length
*	11808	11907: gap of unknown length
*	11908	20464: contig of 8557 bp in length
*	20465	29298: gap of unknown length
*	20565	29298: contig of 8734 bp in length
*	29299	45700: gap of unknown length
*	29399	45700: contig of 16302 bp in length
*	45701	45800: gap of unknown length
*	45801	63662: contig of 17862 bp in length
*	63663	63662: gap of unknown length
*	63763	104507: contig of 40745 bp in length
*	104508	104607: gap of unknown length
*	104608	150643: contig of 46036 bp in length
*	150644	150743: gap of unknown length
*	150744	210923: contig of 60180 bp in length

BASE COUNT	54201	a	51422	c	51731	g	52663	t	906	others
misc_feature			/note="assembly_name:Contig18"							
misc_feature			2530. .4617							
misc_feature			/note="assembly_name:Contig9"							
misc_feature			4718. .11807							
misc_feature			/note="assembly_name:Contig10"							
misc_feature			11908. .20464							
misc_feature			/note="assembly_name:Contig11"							
misc_feature			20555. .29298							
misc_feature			/note="assembly_name:Contig12							
			clone_end:17							
misc_feature			vector_side:right"							
misc_feature			29399. .45700							
misc_feature			/note="assembly_name:Contig13"							
misc_feature			45801. .63662							
misc_feature			/note="assembly_name:Contig14							
			clone_end:SP6							
misc_feature			vector_side:left"							
misc_feature			63763. .104507							
misc_feature			/note="assembly_name:Contig15"							
misc_feature			104608. .150643							
misc_feature			/note="assembly_name:Contig16"							
misc_feature			150744. .210923							
misc_feature			/note="assembly_name:Contig17"							

Query Match	28.7%;	Score 126.6;	DB 39;	Length 210923
Best Local Similarity	67.8%;	Pred. No. 1.6e-12;		

	Matches	177;	Conservative	0;	Mismatches	84;	Indels	0;	Gaps	0;
QY	11	ccgtgaccgccccttccagtgctgctgtgtggaagcgttccgcgacaaagcccaact	70							
Db	98278	CGGGGAGACGGGCCCTACCCGGTCGGAGATGGGCGCAAGGCTTCAACCGAAGATCCACC	98219							
QY	71	tgatgctctaccgcgcgctgacacagcgagcgagcggcccaacagtcgcccagatgcgga	130							
Db	98218	TGGGCTGCACCAAAAGAACCCACACCGGGAGCGGCCCTTCCCTGACGGAATGGAGA	98159							
QY	131	aggcgttaccacaaatagaccctatctgactctgcgacgcygagctacacccgcgaggaagc	190							
Db	98158	AGCGCTTTCGCAAGAGACGGCCTTGATTCGGACACGGCCATTCATGGGCGAGAGGC	98099							
QY	191	ccctacccgctgcaaaagatgctgcgcgcgcttccgcgacaaacccaactctgctctcaca	250							
Db	98098	CCCTACCAATGGCGACAGTGGCGACGCAAGCTTCAACGACGACACTTGTTGGCGCACC	98039							
QY	251	gcaagatcacacaagcatcg	271							
Db	98038	AAAGGTCACCAAGACGGCG	98018							

RESULT	10			
LOCUS	MMZFP29	2094 bp	mRNA	13-APR-1992
DEFINITION	Musculus zfp-29 gene for zinc finger protein.			
ACCESSION	X5126			
VERSION	X5126.1	GI:55470		
KEYWORDS	spermatogenesis; zfp-29 gene; zinc-finger.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
	Rodentia; Sciurognath; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 2094)			
AUTHORS	Denny,P.			
TITLE	Direct Submission			
JOURNAL	Submitted (03-NOV-1990) Denny P., Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, LONDON, SW3 6JB, UK			
REFERENCE	2 (bases 1 to 2094)			
AUTHORS	Denny,P. and Ashworth,A.			
TITLE	A zinc finger protein-encoding gene expressed in the post-meiotid			
JOURNAL	Phase of spermatogenesis			
MEDLINE	Gene 106 (2), 221-227 (1991)			
FEATURES	92039080			
	Location/Qualifiers			
source	I..2094			

misc_feature

higher protein-encoding gene expressed in the post-meiotic spermatogenesis
(2), 221-227 (1991)

Location/Qualifiers

1. 2094

/organism="Mus musculus"

/strain="ssp. domesticus"

/db_xref="taxon:10090"

/dev_stage="adult male"

/tissue_type="testis"

/clone_lib="lambda Zap cDNA"

104. 1948

/gene="Zfp-29"

104. 1948

/gene="Zfp-29"

/codon_start=1

/protein_id="CAA38920.1"

/db_xref="GI:55471"

/translation="MAAEYAVSTPLSLVQVQEDDEAEYTHILLEDNAVQENLV
QEDDEPSEEPQSGSGKSPQEDBAEGPQGALVREELCRMLREYEHTRKQMTLVLD
REIDAMDEHPRPSSEAEVALVLDLQVTRHSDPELTSENGENSEMDEEGVEHGMF
EIPYEGGGGGOOSGDSPEDDCGSGGAQGAEGDEPRVYPSGREGVGLIDQSTYF
PYRCDDCGSEFSSANLITHORLHTEKRYKCEBCKGSEFDSGNSFNRHDTGTEK
SCPGCGSEFNSSANLHOGIHTEKRYPCACGSEFNSNLIHORIHGTEKPYK
TEGCGRQSSALITHRRHTGKPYCCGCECKNSPSNLAHRRHLYKPKCLL
GKGFSSSSSLIAHGHTHTEKRYECLITGSESEFSSNLIKHORHTEKRYKCGDGG
KGSQNSQALVHORHTGKPYKCLMCGSEFSGSLVHQAHLGDGPYKPCGCGG
FSSNSVLIITHORHTEKPYRCECKGSEFSSNSNLIHQRHLKELY"

1952. 1945


```

TITLE      Large Scale Sequence Analysis and Annotation with the Sequence
JOURNAL    Comparison Analysis (SCAN) System
REFERENCE  3 (bases 1 to 154218)
AUTHORS    Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meinke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
            McMurry,K., Han,C. and Deaven,L.
TITLE      Direct Submission
JOURNAL    Submitted (08-JAN-2000) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M886, Los
            Alamos, NM 87545, USA
FEATURES   location/Qualifiers
SOURCE     1..154218
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="RPCL-11.127120"
            /complement(41..330)
            /rpt_family="Alu"
            575..882
            /rpt_family="Alu"
            /rpt_family="Alu"
            1414..1712
            /rpt_family="Alu"
            1595..1862
            /rpt_family="Alu"
            1871..2190
            /rpt_family="Alu"
            2176..2372
            /rpt_family="Al"
            2498..2795
            /rpt_family="Al"
            3187..3504
            /rpt_family="Alu"
            /rpt_family="Alu"
            /complement(3571..3890)
            /rpt_family="Alu"
            /complement(3859..4163)
            /rpt_family="Alu"
            4198..4503
            /rpt_family="Alu"
            4826..5155
            /rpt_family="Alu"
            5749..6045
            /rpt_family="Alu"
            /rpt_family="Alu"
            /complement(6848..6983)
            /note="GRAIL 2 excellent exon, frame 1"
            7473..7776
            /rpt_family="Alu"
            7931..8385
            /rpt_family="Alu"
            8808..9139
            /rpt_family="Alu"
            9394..9682
            /rpt_family="Alu"
            9515..9812
            /rpt_family="Alu"
            9702..9987
            /rpt_family="Alu"
            /rpt_family="Alu"
            9965..10332
            /rpt_family="Alu"
            /complement(10403..10835)
            /rpt_family="Alu"
            /complement(10801..10977)
            /rpt_family="Alu"
            11580..11780
            /rpt_family="MERR42"
            11803..12101
            /rpt_family="Alu"
            12282..12576
            /rpt_family="Alu"
            12643..13073
            /rpt_family="Alu"
            /rpt_family="Alu"

repeat_region 12923..13237
                /rpt_family="Alu"
repeat_region 13420..13725
                /rpt_family="Alu"
repeat_region 13761..14148
                /rpt_family="Alu"
repeat_region 14072..14371
                /rpt_family="Alu"
repeat_region 14401..14401
                /rpt_family="Alu"
                15061..15366
                /rpt_family="Alu"
repeat_region 15383..15706
                /rpt_family="Alu"
repeat_region 16083..16185
                /rpt_family="Al"
                /complement(16192..16347)
                /rpt_family="Alu"
                16559..16831
                /rpt_family="Alu"
repeat_region 168420..18718
                /rpt_family="Alu"
                /complement(18518..18917)
                /rpt_family="Alu"
                /complement(18754..19036)
                /rpt_family="Alu"
                /complement(19456..19559)
                /note="GRAIL 2 excellent exon, frame 1"
                20239..20397
                /rpt_family="MERR20"
                /complement(20808..21104)
                /rpt_family="Alu"
                /complement(22116..22443)
                /rpt_family="Alu"
                /complement(22789..22948)
                /note="GRAIL 2 excellent exon, frame 1"
                23346..23646
                /rpt_family="Alu"
                /complement(24046..24341)
                /rpt_family="Alu"
                25205..25461
                /rpt_family="Alu"
                25605..25862
                /rpt_family="Al"
                25914..25981
                /rpt_family="Alu"
                /complement(26044..26355)
                /rpt_family="Alu"
                28437..28732
                /rpt_family="Alu"
                30597..30772
                /note="GRAIL 2 excellent exon, frame 2"
                31300..31655
                /rpt_family="Alu"
                32059..32124
                /rpt_family="Alu"
                /complement(32302..32492)
                /rpt_family="Alu"
                32507..32839
                /rpt_family="Alu"
                33649..34040
                /rpt_family="Alu"
                /complement(34202..34278)
                /rpt_family="LRR37B"
                34279..34556
                /rpt_family="Alu"
                35017..35495
                /rpt_family="Alu"
                /complement(36639..36991)
                /rpt_family="Alu"
                37667..37808
                /note="GRAIL 2 excellent exon, frame 0"
                /complement(37920..37961)

```

Query Match	27.98;	Score 123.2;	DB 10;	Length 154218;
Best Local Similarity	60.28;	Pred. No. 6.3e-12;		
Matches 242;	Conservative	0;	Mismatches 153;	Indels 7; Gaps 2

OY	8	atgcgcgtcagacgcgcgcctccagtgctgctgtttggaagaagctctgcgcacaagccca	67
Db	45624	ATGGGGCTGACAAAGCGGTACACCTTCCCGGAATTGGCAAAAGCTTTCACCAAGACGTCC	455655
OY	68	actgatctgcctcaccgcgcgctgcaacagggcgagccgccaccagttccccagatgcg	127
Db	45564	ACTTGACCAAGCACACGAGGCGACACACACAGGGGGAGCGGCGCTTACAAAGTCCTAGTCTGG	45505
Db	128	ggaagcgcttaccacaataagccctatctgacttcgacccgacgcataccaacccgagaga	187
Db	45504	GGAAGGCGTTTAGCGACCGCTCCAACTTCAGACGCGCACAGAGGGGTGCACACAGCGGAGA	45445
OY	188	agccctaccgcgtgcacaagaatgctgcgcgcgcttcctgcgcacaacccaactgcgtcttc	247
Db	45444	AGCCCTACCCCGTGCCTCCAGTGTGGGAAGCGCTTTCAGCGACAGGCTCCACCTGGTATCC	45385
OY	248	acagacaatcttcacaacagcatctccgaggggtcgcgccagacgcgcgcgcgcgcggagacc	307
Db	45384	ACCGAGGACACAC-----ACGGGGAGCGGCGCCTATGCTCTGCACCCAGTGGGGGAAGG	45330
OY	308	cccacgtgcagccgcgcgcgcgcagagatccgcgcgcgcgcgcacccccgcgcgtactctga	367
Db	45329	CTTCAACAACAGCTGCGCACTTCAGGCGCCACACGCGCGGACGACACAGGAGGAAGCCCTA	45270
OY	368	aaccgcccacgaagc	409
Db	45269	CACCTGGCCC--GGCTGTGGCCGGGGCTTCCGCGCGGGCACC	45230

RESULT	13	
HSJ59G15/C		
LOCUS	HSJ59G15	
DEFINITION	Homo sapiens chromosome 1 clone RP4-59G15 map p12-13.2, ***	06-OCT-2000

```
ACCESSION      SEQUENCING IN PROGRESS ***, 6 unordered pieces.
VERSION        ALI09966.32   GI:9801558
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE          1 (bases 1 to 138278)
JOURNAL        Direct Submission
               Submitted (04-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
               Requests: clonerequests@sanger.ac.uk
               On Aug 14, 2000 this sequence version replaced gi:9588436.
COMMENT        ----- Genome Center
               Center: Sanger Centre
               Center code: SC
               Web site: http://www.sanger.ac.uk
               Contact: humquerry@sanger.ac.uk
               Project Information
               Center project name: dj599G15
               ----- Summary Statistics
               Assembly program: XGAP4; version 4.5
               Sequencing vector: M13; M77815; 25% of reads
               Sequencing vector: plasmid; L08752; 74% of reads
               Chemistry: Dye-terminator ET-amersham; 0% of reads
               Chemistry: Dye-terminator ABI; 1% of reads
               Chemistry: Dye-terminator Big Dye; 94% of reads
               Chemistry: Dye-primer-amersham; 2% of reads
               Chemistry: Dye-primer Big Dye; 1% of reads
               Consensus quality: 136114 bases at least Q40
               Consensus quality: 137176 bases at least Q30
               Consensus quality: 137544 bases at least Q20
               Insert size: 137778; sum-of-contigs
               Insert size: 126614; 2.3% error; agarose-fp
               Quality coverage: 6.30x in Q20 bases; sum-of-contigs Quality
               coverage: 8.38x in Q20 bases; agarose-fp
               -----
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 6 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *
               * 1. 8020: contig of 8020 bp in length
               *    8021 8120: gap of 100 bp
               *    8121 15996: contig of 7876 bp in length
               *    15997 16096: gap of 100 bp
               *    16097 19055: contig of 2959 bp in length
               *    19056 19155: gap of 100 bp
               *    19156 22087: contig of 2932 bp in length
               *    22088 22187: gap of 100 bp
               *    22188 29998: contig of 7811 bp in length
               *    29999 30098: gap of 100 bp
               *    30099 138278: contig of 108180 bp in length.
               Location/Qualifiers
                   1..138278
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p12-13.2"
                     /clone="RP4-599G15"
                     /clone_1lb="RPC1-4"
                   1..8020
                     /note="assembly fragment:01125
                     fragment chain:1"
                   8121..15996
                     /note="assembly fragment:02016
                     fragment chain:1"
                   16097..19055
                     misc_feature
                         misc_feature
                             misc_feature
```


THIS PAGE BLANK (USPTO)